

SPOTTING THE SYNTHETIC

From detection to discernment

Pastoral media discernment for seminarians

Test everything; hold fast what is good · 1 Thessalonians 5:21



Real, synthetic, or unsure?

Vote first. Explain the evidence second.



THE STRONGER QUESTION

What exactly is being claimed?

What evidence establishes that claim?

Authenticity, context, intent and evidence

Course outcomes

Use the deck for a 90-minute session, a two-hour workshop, or independent study

Name the disorder

Distinguish misinformation, disinformation, malinformation, miscontext, cheapfakes, deepfakes and synthetic media.

Explain generation

Describe how systems create text, images, video, voices and synthetic personas.

Separate the layers

Distinguish the media file, its context and the attached claim.

Apply PASTOR

Pause, investigate and respond before believing, sharing or correcting.

Use tools in order

Begin with direct confirmation, then search, provenance checks and specialist tools.

Respond pastorally

Protect dignity while addressing scams, rumours, impersonation and harmful media.

Truth as a pastoral responsibility

False information injures real people

Reputation and dignity

A short clip or false accusation can trigger summary judgement.

Safety and peace

Rumours can provoke hostility, communal tension or retaliation.

Money and access

Impersonation can steal funds, credentials or confidential information.

Trust in the Church

Uncritical forwarding weakens confidence in pastors and institutions.



Christus Vivit

Digital participation requires pastoral prudence

Promise and risk

The digital world supports participation, but also manipulation, exploitation, violence and isolation.

Closed information circuits

Like-minded groups can reinforce fake news, prejudice and hatred.

Online summary trials

A person's reputation can be damaged before facts and context are established.

Seminarian's discipline

Resist passive consumption and impulsive judgement. Seek the full recording and context.



Artificial intelligence and wisdom of the heart

World Day of Social Communications 2024

Cognitive pollution

Partly or completely false narratives circulate as though they were true.

Deepfake persons

Images and cloned voices can make someone appear to say what was never said.

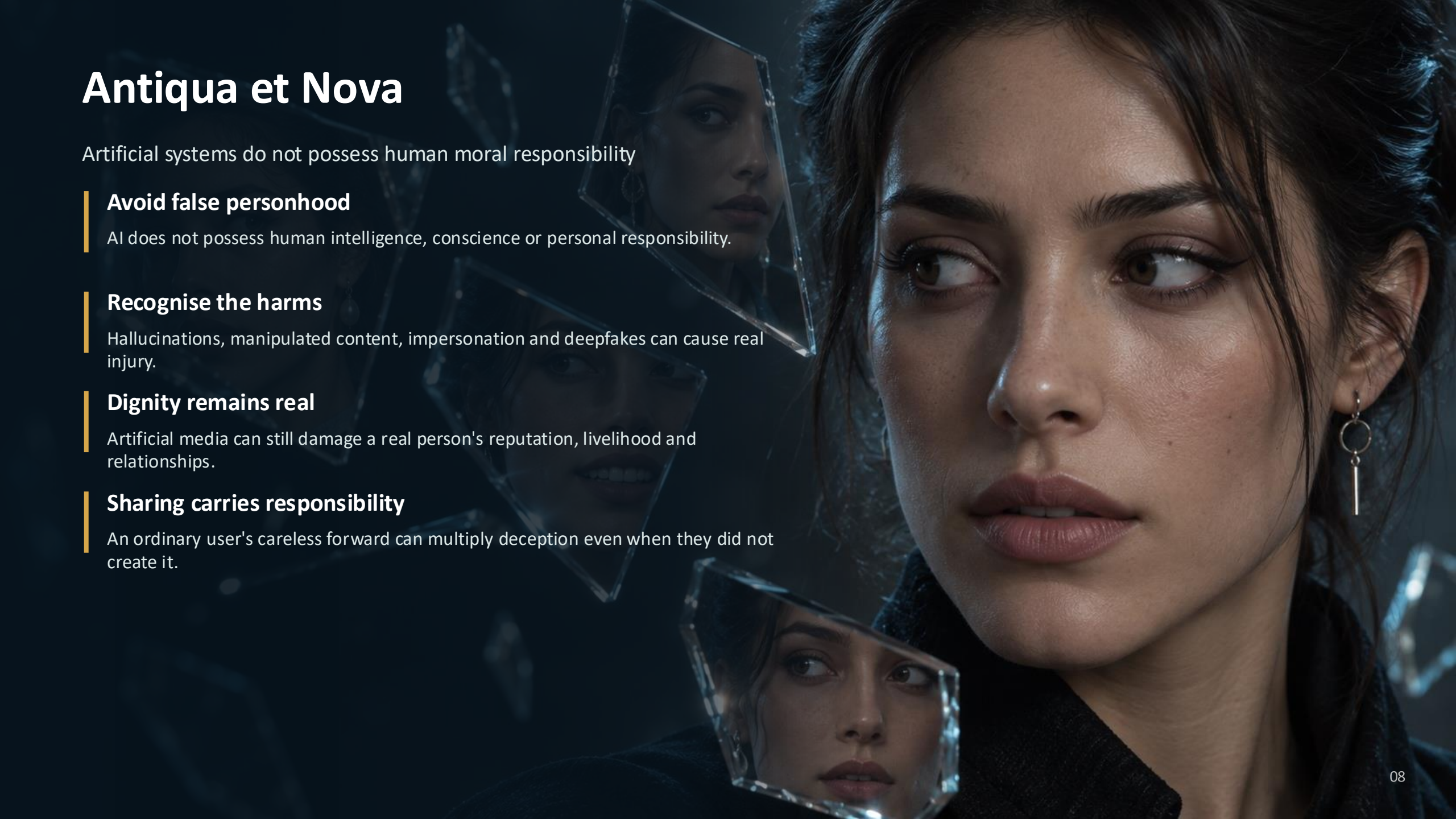
Simulation and distortion

Simulation becomes perverse when it distorts our relationship with people and reality.

Algorithms carry choices

They can serve knowledge and encounter, or profit, polarisation and groupthink.

Antiqua et Nova



Artificial systems do not possess human moral responsibility

Avoid false personhood

AI does not possess human intelligence, conscience or personal responsibility.

Recognise the harms

Hallucinations, manipulated content, impersonation and deepfakes can cause real injury.

Dignity remains real

Artificial media can still damage a real person's reputation, livelihood and relationships.

Sharing carries responsibility

An ordinary user's careless forward can multiply deception even when they did not create it.

Magnifica Humanitas

Truth belongs to the common good

Technical claims age quickly

Scientific study must continue alongside moral and spiritual discernment.

Public truth needs method

Verification, cross-checking and responsible argument support a shared search for truth.

Trust grows relationally

Honest exchange and transparent communication strengthen community.

Education is essential

Christian communities should cultivate critical thought and responsible digital practice.



A pastoral synthesis

Truth is a common good

The loudest voice does not own it.

Dignity comes before virality

Do not enlarge a victim's humiliation to prove a post false.

Prudence slows judgement

Wait until adequate evidence is available.

Charity corrects gently

Distinguish a mistaken sharer from a deliberate deceiver.

Justice acts promptly

Fraud, harassment, exploitation and incitement require timely action.

Information disorder

'Fake news' hides important differences

Method matters

An honest mistake differs from a deliberate deception or weaponised truth.

Context matters

Real media can mislead when date, place, identity or sequence changes.

Intent matters

Humour, manipulation, fraud and incitement call for different responses.

Evidence remains essential

A category helps diagnose the method, but does not prove what happened.



Error, deception and weaponised truth

MISINFORMATION

False or misleading information shared without intent to deceive. Example: an old flood image forwarded as today's event.

DISINFORMATION

False information deliberately created or distributed to deceive or harm. Example: a fabricated diocesan cancellation notice.

MALINFORMATION

Reality used out of context or weaponised to harm. Example: a private pastoral message leaked to humiliate its sender.

Context and deceptive editing

Authentic material can carry a false meaning

MISCONTEXTUALISED MEDIA

Authentic media receives a false date, place, person or explanation. Example: a foreign incident presented as local.

CHEAPFAKE

Real media is deceptively cropped, slowed, reordered, captioned or selectively edited without advanced AI.

Common pastoral form

A homily stops just before the sentence that explains its meaning.



Synthetic, automated and performative media

DEEPPFAKE

A person's face, body or voice is created or altered with machine-learning techniques.

AI SLOP OR SYNTHETIC PERSONA

Low-accountability synthetic content or fabricated identities pursue reach, money or influence.

FULLY SYNTHETIC MEDIA

The people, place, event, image, sound or video may never have been recorded.

SATIRE OR PARODY

Manipulated media used for humour or criticism without presenting itself as factual reporting.

Categories can overlap

One incident may involve several methods and harms



Describe the method precisely. Then state what the evidence proves.

Three layers of every claim

Separate the file from what people say about it



MEDIA FILE

Recorded, edited, generated, or mixed



CONTEXT

Date, place, identity, sequence



CLAIM

What the sender says happened

An authentic file can carry a false caption. Synthetic media can be labelled and used responsibly.

How generative media is produced

Systems learn patterns and generate new outputs from instructions

Training patterns

Models learn statistical relationships from very large collections of data.

Prompted generation

Instructions guide text, images, audio or video toward a requested result.

Mixed media

A single clip can combine recorded, edited and generated elements.

Fluency without witness

Convincing output does not guarantee accuracy, truth or moral responsibility.



Text generation

Fluent language can still contain invented facts

Prediction

A language model estimates which words or tokens are likely to follow the prompt and prior text.

Imitation

It can reproduce tone, structure and style without possessing the speaker's experience or intention.

Hallucination

A confident quotation, reference, name or explanation may be inaccurate or invented.

Verification

Trace important claims, quotations and citations to reliable originals.

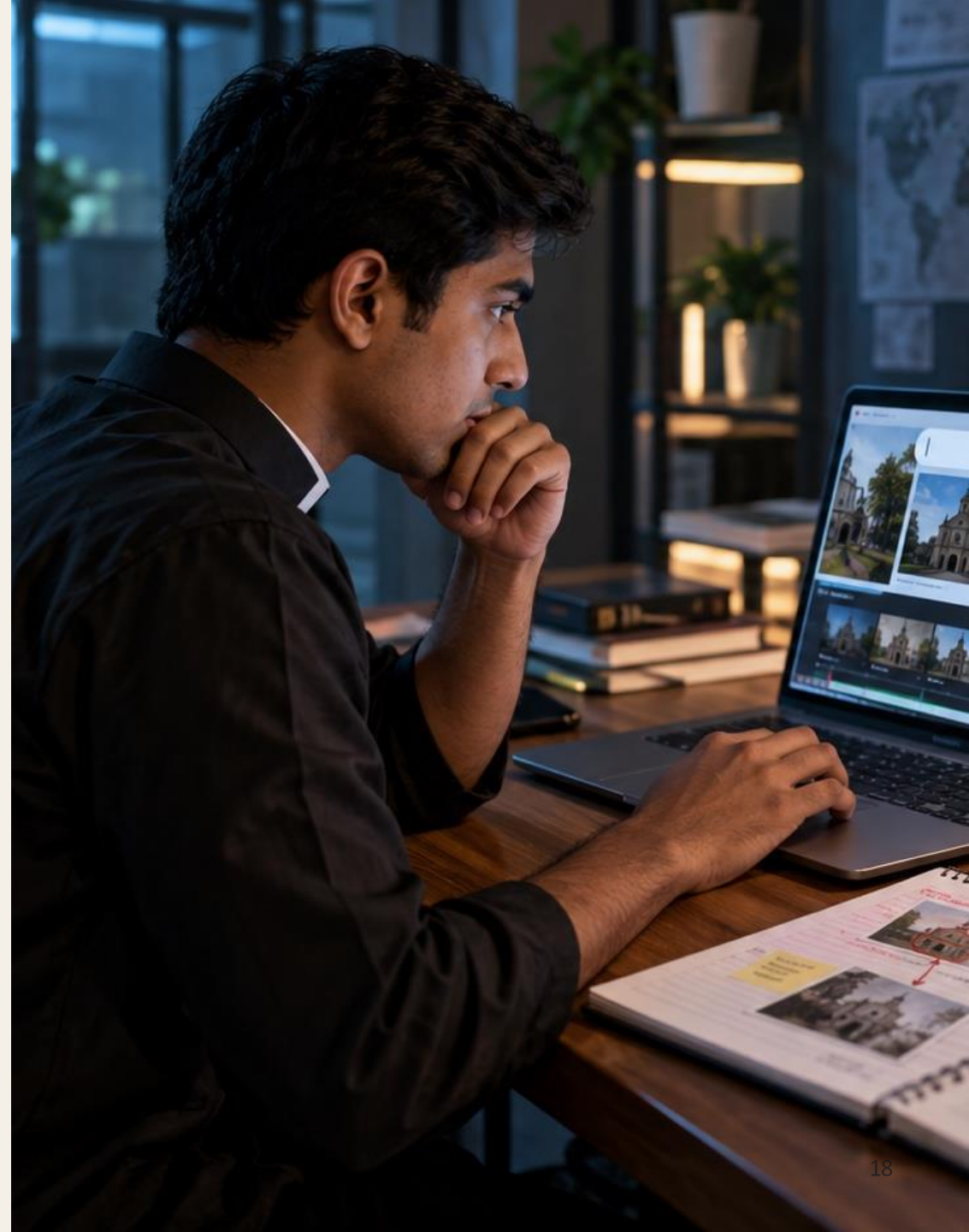


Image and video generation

Generation, animation and editing can coexist in one clip

Diffusion

Many image systems refine visual noise until it matches the prompt.

Temporal continuity

Video systems try to maintain identity, motion, lighting and objects across frames.

Common operations

Animate a still, replace a face, alter expression, synchronise lips, remove objects or build a composite scene.

Hybrid media

A clip may contain authentic footage, ordinary editing and generated elements together.

Voice cloning and synthetic audio

A familiar voice no longer proves identity

Learned characteristics

Systems model pitch, rhythm, pronunciation and timbre from voice samples.

New sentences

The model can speak words the person never recorded.

Hidden weaknesses

Compression, background noise and short urgent messages can mask defects.

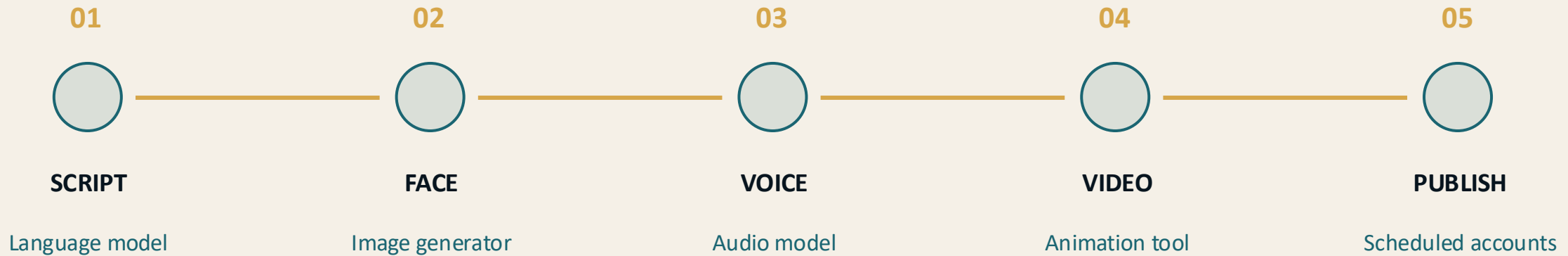
Safer identity check

Confirm through a known number or another independent channel.



Synthetic personas and production pipelines

The strongest clue may appear in the account history



Check repetitive posting, missing identity, copied descriptions, strange commercial links and the absence of an independent trace.

Two essential distinctions

Ethical judgement includes disclosure, consent, dignity and truth

AI-GENERATED

does not automatically mean false

A labelled reconstruction, illustration, translation or accessibility aid may communicate responsibly.

AUTHENTIC

does not automatically mean trustworthy

A real photograph or recording may be old, staged, selectively edited or described falsely.

Emotion and urgency

The pause restores freedom

Narrowed attention

Anger, fear, disgust, wonder and pity can reduce careful scrutiny.

Urgency as pressure

'Share before this is deleted' tries to turn anxiety into obedience.

Emotion is not proof

A strong reaction does not establish that the claim is false or true.

Spiritual discipline

Name the emotion and slow down before acting.



Familiarity and repetition

Count independent evidence, not reposts

Repeated claims begin to feel true

Familiarity can be mistaken for evidence.

Many accounts may share one source

Apparent confirmation can come from copying rather than independent reporting.

Trace the chain backward

Ten posts pointing to one unverified clip still provide one unverified source.

Ask what each source independently knows

Separate eyewitness evidence from recycled commentary.

Authority and relationship

Impersonators borrow trust

Relational authority

A bishop, priest, superior, family member or institution can prompt immediate compliance.

Sensitive requests

Money, credentials, secrecy and public statements require stronger confirmation.

Separate channel

Contact the person through a number or directory already known to you.

Courtesy supports prudence

A genuine leader will understand a careful call-back.

Confirmation and group identity

Apply the same standard to welcome and unwelcome claims

Flattering claims bypass scrutiny

Content that praises our group can feel safe before it is checked.

Hostile claims simplify a rival

A dramatic accusation may satisfy prior suspicion.

Evidence standards should remain stable

Require the same quality of proof from allies and opponents.

Disagreement does not prove deception

Investigate the claim rather than the identity of the group sharing it.

The liar's dividend

Deepfakes also make authentic evidence easier to deny

Second-order danger

Awareness of synthetic media allows a guilty person to dismiss real evidence as fake.

Two extremes

Avoid believing every vivid recording and rejecting every inconvenient one.

Preserve originals

Maintain the best available file, source, time and chain of custody.

State certainty honestly

Use provenance and independent evidence; acknowledge unresolved questions.



From detection to discernment?

Pixel inspection comes after source and context

1

CLAIM

What is alleged?

2

SOURCE

Who first published it?

3

CONTEXT

When and where?

4

CORROBORATION

Who independently confirms it?

5

MEDIA

What clues and provenance remain?

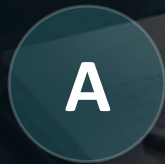
Visual anomalies begin an investigation. They rarely complete it.

The PASTOR verification method

A memorable workflow for responsible judgement



Pause



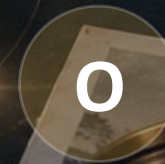
Ask



Search



Trace



Observe



Respond

P Pause

Interrupt the automatic action

Do not forward, accuse or pay

Stop before the requested action creates harm.

Read the exact wording

Notice urgency, secrecy and the consequences of being wrong.

Write one neutral claim

Separate what the media shows from what the caption says.

Choose 'unsure' when needed

Insufficient evidence does not require a forced verdict.

A Ask about the source

The forwarder may not be the creator

Who first recorded, wrote or uploaded it?

Trace beyond the account that sent it to you.

Does the account have a credible history?

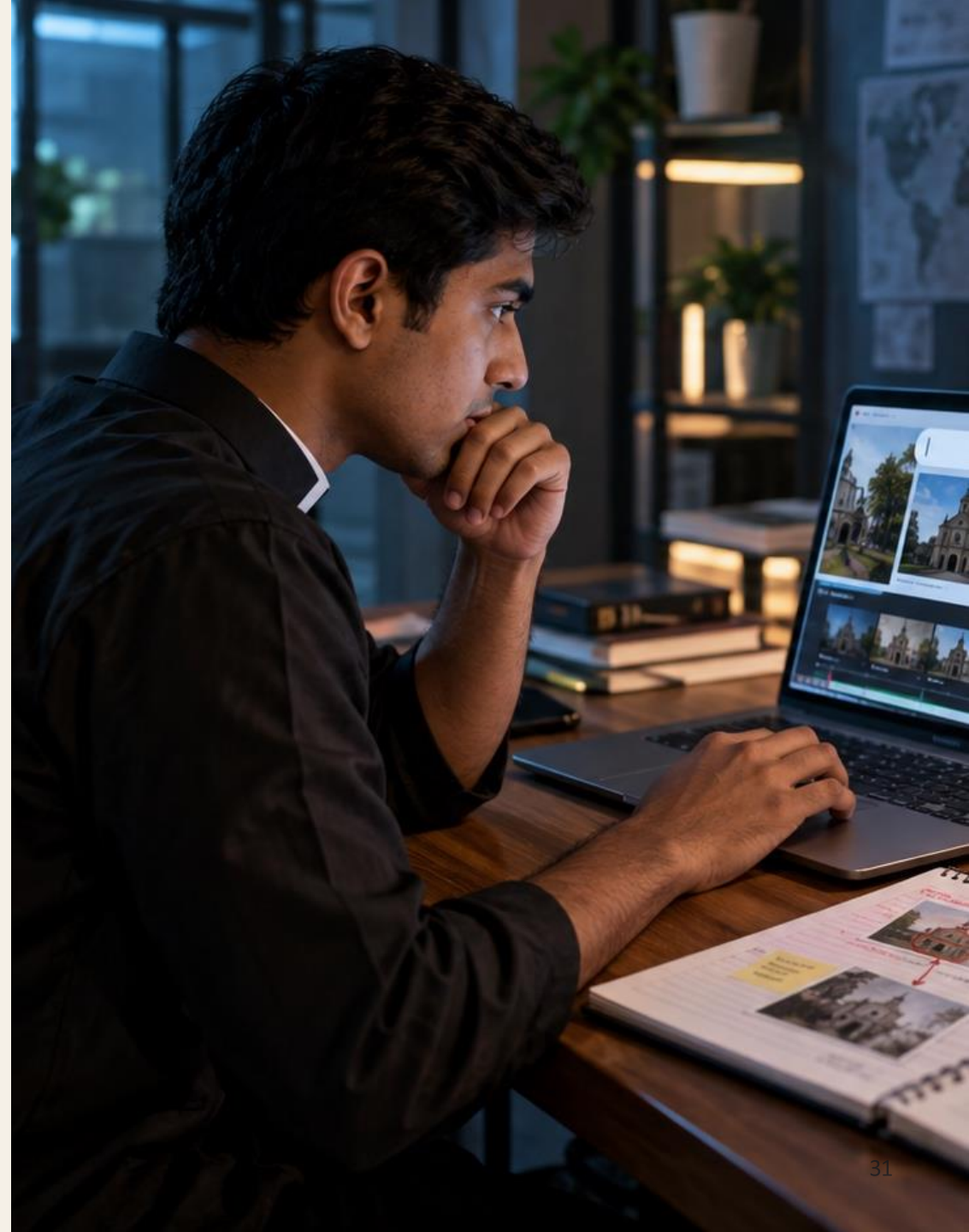
Review older posts, identity, corrections and institutional links.

Is the website independently confirmed?

Find it through an official directory or known institutional channel.

What action is requested?

Secrecy, money, credentials and unusual favours raise the verification threshold.



S Search laterally

Leave the post and investigate elsewhere

Search exact language

Put a distinctive sentence or quotation in quotation marks.

Add investigative terms

Search the name with transcript, clarification, fraud or fact check.

Use independent sources

Avoid counting copies of one original as separate evidence.

Prefer authoritative originals

Use official Church, government and credible local sources where appropriate.

T Trace to the original

Reposts strip away date, credit, quality and context

Find the earliest available upload

Work backward through links and reposts.

Prefer the longest version

Locate the full homily, interview, video or document.

Compare captions and dates

Note what changed, disappeared or was added.

Preserve evidence

Keep the URL and a screenshot when reporting may be necessary.



O Observe the evidence

Treat internal anomalies as clues

Compare across time

Look for changes in identity, objects, lighting, reflections and continuity.

Use tools purposefully

Reverse search, keyframes, metadata and Content Credentials should answer a specific question.

Check known reality

Compare speech, setting, chronology and independent recordings.

Combine evidence

A strong conclusion normally rests on several forms of support.

R Respond responsibly

Uncertainty does not require a public accusation

VERIFY

Seek stronger evidence or direct confirmation.

WITHHOLD

Do not share while important facts remain unresolved.

CORRECT

Publish verified facts and accurate context.

REPORT

Preserve evidence and use platform, institutional or legal channels.

ACCOMPANY

Protect and support the person harmed by fraud, humiliation or abuse.

Inspection without false confidence

The same symptom may have innocent and deceptive causes

Platforms re-encode media

Compression changes edges, motion, colour and sound.

Phones apply computational effects

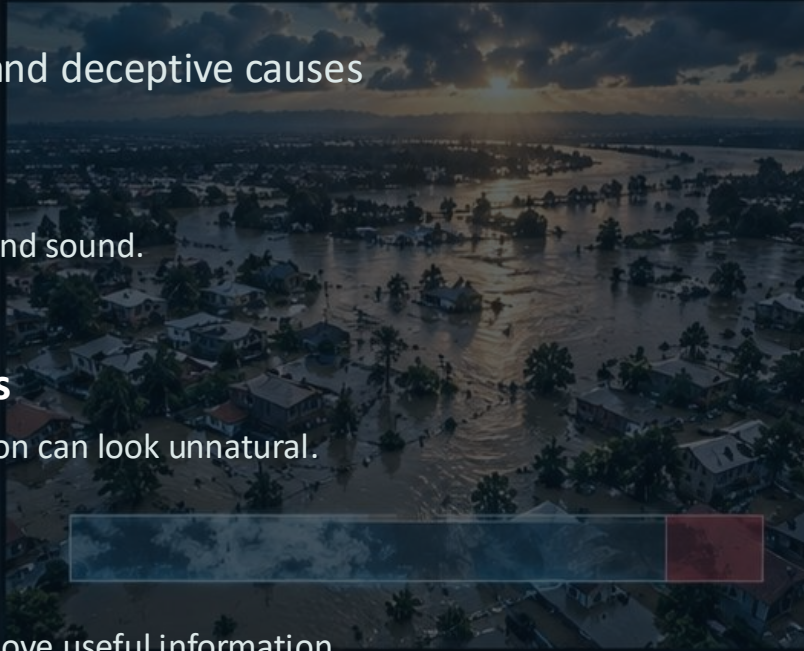
Stabilisation, portrait mode and noise reduction can look unnatural.

Ordinary editing alters evidence

Cropping, filters and repeated forwarding remove useful information.

High-quality fakes may look flawless

The absence of anomalies does not establish authenticity.



Inspection suggests questions. It rarely establishes the whole truth.

Possible visual clues

Continuity and interaction

Temporal inconsistency

A face, object, garment, logo or background changes shape across frames.

Interaction errors

Hands pass through objects, contact points slip or objects appear without a plausible cause.

Identity drift

Facial features, age, hairline, teeth or accessories shift during movement.

Biological or physical impossibility

Motion, anatomy, gravity or cause and effect becomes inconsistent.



Possible visual clues

Lighting, text and boundaries

Lighting and reflection conflicts

Shadows, highlights, mirrors or eyeglasses do not agree with the scene.

Text and symbol instability

Writing changes across frames or remains semantically meaningless.

Boundary artefacts

Hair, ears, jewellery or face edges blend unnaturally during movement.

Innocent lookalikes

Low frame rate, motion blur, beauty filters, weak connections and edited backgrounds can create similar effects.

Possible audio clues

Confirm identity through another channel

Cadence and emphasis

Delivery does not fit the speaker's normal manner or the sentence meaning.

Acoustic discontinuity

Breathing, mouth sounds, room echo or background noise changes abruptly.

Pronunciation

Names, local words or emotional expressions sound unusual.

Audio-image mismatch

Lip movement drifts after a cut, or the voice stays unnaturally clean in a noisy scene.



Documents and quotation cards

Trace the words beyond the graphic

Search a distinctive sentence

Use quotation marks and look for the complete speech, homily or interview.

Check institutional details

Review issuing office, date, reference number, signature block and distribution channel.

Compare known formats

Look at verified documents from the same institution.

Watch for contradictions

Invented citations, impossible dates, wrong office titles and false domains are warning signs.



Weak rules that create false confidence

These clues can begin an investigation, but cannot end one

✗	VIDEO UNDER 15 SECONDS	Real clips are cropped; synthetic clips can be extended.	Trace the full clip and confirm the event.
✗	HANDS LOOK CORRECT	Current systems often render hands well; blur can distort real hands.	Seek several clues and provenance.
✗	THROAT DOES NOT MOVE	Angle, anatomy, clothing, frame rate and compression make this unreliable.	Contact the speaker and find the original.

More weak rules

A detector score remains one item of evidence

×	MUFFLED AUDIO	Compression and poor microphones commonly cause it.	Confirm identity through a trusted channel.
×	ONE CAMERA ANGLE	Many real events have only one surviving recording.	Search for independent evidence beyond the clip.
×	DETECTOR SAYS 99%	Detectors can misclassify and fail on unfamiliar generators.	Ask what other evidence supports the score.

The investigator's toolkit

Begin with the strongest and fastest evidence

1

DIRECT CONFIRMATION

Known number · ordinary search · official source

2

REVERSE SEARCH

Google Lens · video keyframes · earlier uses

3

PROVENANCE

Content Credentials · EXIF · file history

4

ACCOUNT AND DOMAIN

Posting history · identity · registration

5

AI DETECTORS

Supporting evidence only

Presence of strong evidence adds confidence. Absence of provenance proves nothing by itself.

Level 1 Direct confirmation

For money, credentials, secrecy or urgent action

End the incoming interaction

Do not continue under the sender's pressure.

Use a known number

Call the person or office through a saved contact or confirmed directory.

Shared verification phrase

Use a private phrase or two-person approval for exceptional financial requests.

Preserve the suspicious message

Keep the number, account, time and requested action for reporting.

Level 1 Search and official sources

Open the original source, not only the search snippet

Ordinary web search

Search exact quotations, names, dates, places and document titles.

Church sources

Use the Holy See, bishops' conference, diocesan website or confirmed official channel.

Local confirmation

Consult credible local reporting and people who directly know the event.

Government claims

PIB Fact Check covers claims about Union Government bodies and policies.



Level 2 Google Lens

Trace an image's web history

Upload the image or screenshot

Use the clearest available version.

Crop to the distinctive subject

A building, person, vehicle or landmark can improve matching.

Review earlier dates and captions

Look for the same picture used in another place or event.

Interpret carefully

Reverse search finds related uses; it does not automatically identify the truthful caption.



Level 2 InVID WeVerify

Extract keyframes and reverse-search a video

01 Copy the public URL or save a lawful working copy

02 Generate representative keyframes

03 Choose two or three distinctive frames

04 Search through more than one engine when possible

05 Compare earliest result, caption, date, place and length

06 Record the URLs and what each result proves

Level 3 Provenance and metadata

Origin evidence and claim truth are different questions

Content Credentials

C2PA can provide signed information about creation, signer and declared edit history.

Absence proves nothing

Many cameras, apps and platforms do not create or preserve credentials.

Ordinary metadata

EXIF may contain date, device, location or software information.

Supporting evidence

Metadata can be stripped or edited. A genuine file can still carry a false caption.



Level 4 Account and domain investigation

Identity leaves a history

Review the posting timeline

Do older posts support the claimed identity and expertise?

Look for abrupt changes

Topic, language, profile name or audience shifts may indicate a hijacked account.

Compare official links

Confirm the account through the institution's website or public directory.

Check the domain carefully

ICANN Lookup can help, but privacy services and new legitimate sites limit certainty.

Treat popularity as weak evidence

Followers, polished branding and frequent posting can be purchased or automated.

Level 5 AI detectors

Useful for triage, unsafe as a judge

What detectors estimate

They compare a file with patterns associated with known generation or manipulation methods.

Why scores fail

Cropping, recompression, filters and unfamiliar generators can change the result.

How to record the output

Treat the score as one piece of evidence and record the tool, file version and limitations.

Classroom demonstration

Compare an original photo, a compressed screenshot, a generated image and an edited repost.

The evidence log

A serious investigation should be reproducible

Claim and receipt

Write the exact claim, URL, account and time received.

Best available file

Preserve the original or highest-quality lawful copy.

Search and source trail

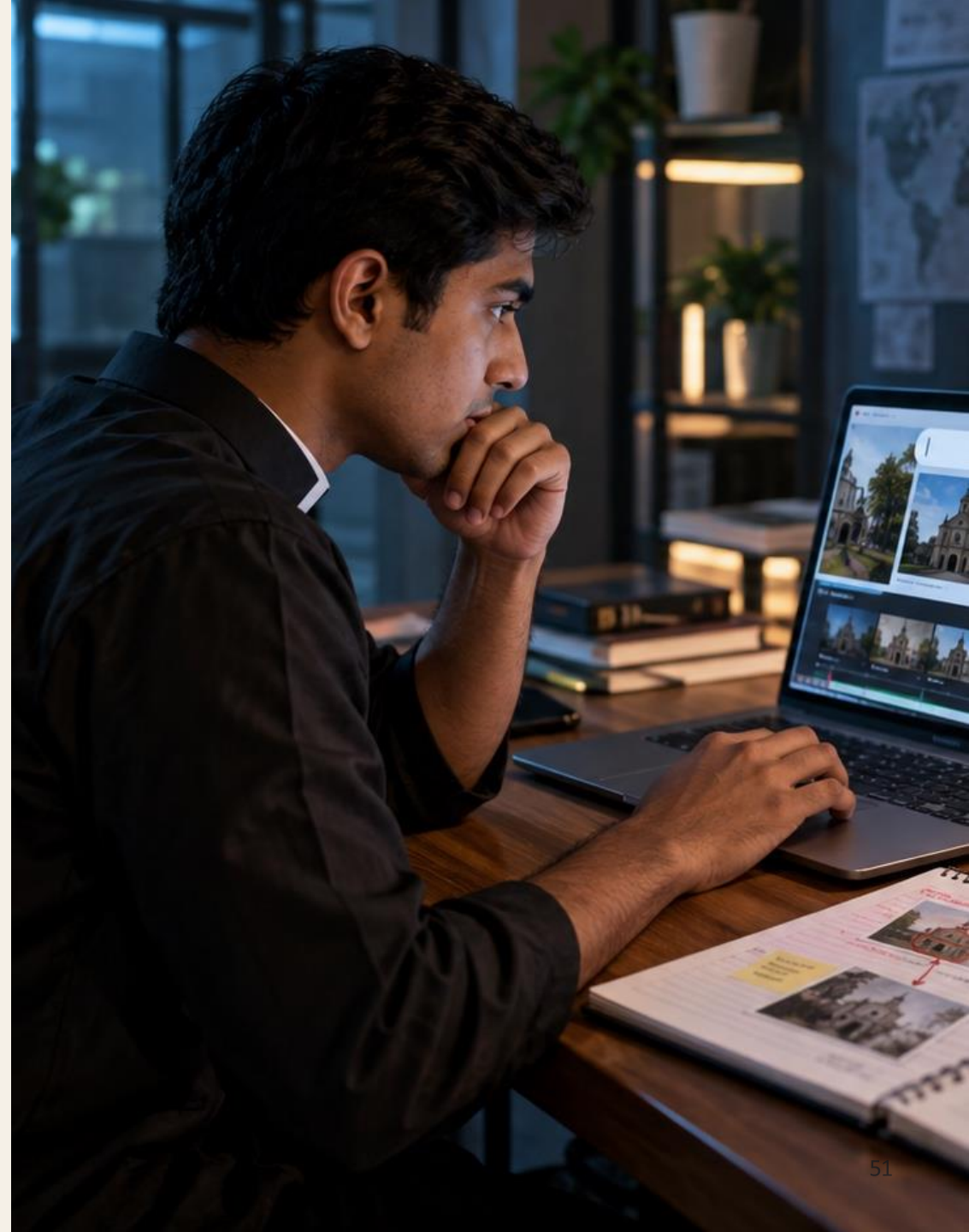
Record terms used, sources consulted and earliest version found.

Tool results and limitations

State what each result supports and what it cannot establish.

Conclusion

Include confidence level, people contacted and unresolved questions.



Pastoral case Urgent transfer

The voice sounds familiar. The request is confidential.

Pause

Make no payment and disclose no credentials.

Check the number

Ask whether it is already known and confirmed.

Call back independently

Contact the bishop's office or authorised person through a saved number.

Preserve and report

Keep the message and account details. Alert likely targets without circulating the cloned voice.

Pastoral case Inflammatory statement

Synthetic, selectively edited, or authentic?

Write the exact alleged words

Avoid relying on the angry caption alone.

Find the first uploader and full event

Search for the complete recording, transcript and other camera angles.

Contact the person or institution

Do this before public defence or condemnation.

Inspect as support

Cuts, audio continuity and lip synchronisation supplement the source investigation.

Pastoral case Puzzling CCTV

Protect faith from gullibility and contempt

Obtain the original file

Do not analyse only a phone recording of a screen.

Reconstruct the setup

Confirm date, camera position, frame rate, maintenance and access.

Test ordinary causes

Consider compression, dropped frames, insects, reflections, camera movement and editing.

Consult competent help

Do not make a theological claim that exceeds the evidence.

Pastoral case ‘The Pope or bishop says’

A beautiful quotation card still needs an original

Search a distinctive phrase

Use quotation marks and the name of the alleged speaker.

Search official documents

Locate the complete speech, homily, interview or statement.

Check alteration

The card may paraphrase, mistranslate or join separate sentences.

Correct with the actual link

If no source is found, call it unverified rather than inventing one.



Pastoral case Communal rumour

Verification becomes a matter of public safety

Do not forward it as a question

Asking 'is this true?' can spread the rumour further.

Confirm on the ground

Contact verified authorities and trusted local leaders.

Investigate location clues

Compare landmarks, weather, language, vehicle plates and earlier uploads.

Communicate verified facts

Discourage retaliation and name one recognised source for updates.



Pastoral case Sexualised deepfake

The immediate priority is safety and dignity

Do not request or circulate copies

Informal analysis can deepen the abuse.

Preserve necessary evidence carefully

Restrict access and protect the person's privacy.

Support reporting and professional help

Assist with platform, cybercrime, legal and safeguarding channels.

Avoid victim-blaming language

The manipulated image is abuse, not the victim's fault.



A clear correction pattern

Replace error with a believable account of the evidence

1

VERIFIED FACT

2

BRIEF FALSE CLAIM

3

METHOD AND EVIDENCE

4

ACCURATE CONTEXT

5

CLEAR ACTION

Lead with what is true. Explain how you know. Tell people what to do next.

Correct the claim. Protect the person.

Truth and dignity belong together

Begin with a shared value

Concern for community, safety, faith or truth can lower defensiveness.

Avoid public ridicule

A mistaken sharer differs from a person deliberately causing harm.

Limit exposure

Describe, blur or restrict access rather than embedding the harmful media everywhere.

Seek consent

Do not use a victim's story in training without permission.

Keep legitimate investigation open

False media must not become an excuse to silence real complaints or uncomfortable facts.



Parish and diocesan readiness

Prepare before the crisis

Name responsible persons

Decide who receives reports and who may speak publicly.

Maintain official directories

Keep confirmed phone numbers, websites and communication channels updated.

Require second confirmation

Exceptional financial requests need two-person approval or a verification phrase.

Train likely targets

Clergy, staff and volunteers should recognise impersonation and reporting routes.

Parish and diocesan incident response

Preserve, assess, confirm, communicate and learn

Preserve the evidence

Keep the original message, URL, time and account details before they disappear.

Assess immediate risk

Consider safety, finance, reputation and communal peace.

Confirm through trusted channels

Contact the impersonated person and relevant authorities.

Issue one concise correction

Use the recognised official account and evidence-based wording.

Review the incident

Update training, procedures and technical safeguards.

Practical investigation exercise

The goal is a transparent conclusion, not a lucky guess

ROUND 1

2 minutes

Choose authentic, synthetic or unsure.
Record the emotion and visual detail that shaped the first impression.

ROUND 2

12–15 minutes

Write the claim, find the source, conduct two lateral searches, trace the original and use one suitable tool.

ROUND 3

5 minutes

Prepare a 100-word parish response that states verified facts, protects dignity and gives a clear action.

Evaluation questions

Judge the quality of the investigation, not only the final label

Did the group investigate the claim rather than only the pixels?

Source and context should guide media inspection.

Did it use independent sources?

Copies of one post do not provide corroboration.

Did it distinguish evidence from inference?

Participants should say what a result proves and what it merely suggests.

Did it allow an unverified conclusion?

Insufficient evidence should remain insufficient.

Would the response reduce harm?

The correction should protect dignity and help people act responsibly.

A 90-minute session route

Short inputs, frequent decisions, one live investigation

0–8

Challenge

8–16

Church teaching

16–28

Information disorder

28–38

Generation

38–50

Detection limits

50–62

PASTOR

62–78

Live investigation

78–90

Pastoral response

End with one personal commitment: verify before amplifying.

A two-hour practical workshop

Retain the same opening, then deepen practice

30 minutes Small-group investigation

Each group follows PASTOR and keeps an evidence log.

15 minutes Findings

Groups state what they know, how they know it and what remains unresolved.

15 minutes Response simulation

Use an impersonation scam, inflammatory clip or communal rumour.

Closing decision

Each group names one parish or diocesan procedure that should change.

Curated video use

Show 30–90 second extracts, pause before the reveal, then investigate

NISCORT How to Debunk Fake News

Fact-checking foundations and source verification.

DW News Deepfakes Flooding Your Feed

Disinformation, social trust and the liar's dividend.

10 Tampa Bay How to Spot Deep Fakes

Introductory clues followed by a discussion of why clues evolve.

TODAY What Is Real and AI-Generated

Interactive authentic, synthetic or unsure voting.

Curated video use

Return every clip to evidence and the PASTOR workflow

NOVA PBS How to Detect Deepfakes

Scientific concepts and limits of human detection.

Additional clip one

Use as a case study and verify its central claim with PASTOR.

Corridor Crew Fake YouTube Channels

Synthetic personas, production pipelines and monetised deception.

Additional clip two

Require an evidence log before the class discusses the verdict.

VERIFY BEFORE YOU AMPLIFY

P Pause

A Ask about the source

S Search laterally

T Trace to the original

O Observe the evidence

R Respond responsibly

'I do not know yet' remains a responsible conclusion.

1 Thessalonians 5:21