

Detecting Synthetically Generated Text Documents

Preamble

Language models such as GPT-4, Claude, Gemini, Llama and Mistral have been effectively utilised by adversaries in their information operations for generating press releases, intelligence reports, policy papers and even statements attributed to senior officials. This detection problem requires the use of technical tools combined with proper analysis of the source.

Ten-Point Advisory

1. Identify Stylistic Over-Coherence and Absence of Voice.

The LLM output has an extreme degree of coherence i.e. all sentences are grammatically correct, transitions between paragraphs flow seamlessly and the entire text has a consistent tonal register. Genuine human writing, especially less formal journalism and activism writing, will have phrases typical of human writing, grammatical simplifications, changes in tone and inconsistencies in style which are not often reproduced accurately by AI language model.

POINT 1 – Stylistic Over-Coherence

AI-GENERATED TEXT	VS	AUTHENTIC HUMAN TEXT
✗ Perfect sentence structure throughout ✗ Smooth, uniform paragraph transitions ✗ No idiomatic expressions or shortcuts ✗ Consistent formal register, zero variation		✓ Occasional grammatical shortcuts ✓ Tonal shifts and stylistic quirks ✓ Idiomatic, colloquial phrasing ✓ Inconsistencies that reflect real voice

2. Look for Excessive Hedging and Vague Attribution.

AI disinformation disguised as intelligence or journalistic output will include phrases such as “It is widely believed that...”, “According to many experts...” or “Sources have indicated...” without any named sources. Real intelligence analysis and journalism always name their sources. The consistent omission of named sources is a clear sign of LLM origin.

POINT 2 – Excessive Hedging & Vague Attribution

AI DISINFORMATION SIGNATURE PHRASES — No Named Sources	
"it is widely believed that..."	"many experts suggest..."
"according to multiple sources..."	"sources indicate..."
"analysts say..."	✓ LEGITIMATE DOCUMENTS CITE: Named officials • Specific units Verifiable dates • Published reports

3. Scrutinise Named Entity and Factual Consistency.

AI text generators suffer from a tendency towards hallucination, which entails confidently generating details that are untrue. Every single individual mentioned in any suspected text should be separately verified: units should be compared against official ORBATs; personnel against military or police databases; events against verified chronologies; and casualty numbers against official MOD and PIB reports. If multiple specific facts cannot be verified in an individual document, it is AI-generated.

POINT 3 – Named Entity & Factual Consistency Check	
ENTITY TO VERIFY	CROSS-CHECK AGAINST
Battalion / Unit Numbers	→ Published ORBAT data (official order of battle records)
Officer Names & Ranks	→ Official Army records and gazette notifications
Dates & Event Timelines	→ Verified event logs and official press releases
Casualty Figures	→ MOD / PIB official releases — reject unverified numbers
⚠ Multiple failed cross-checks = HIGH PROBABILITY of AI-generated fabrication — escalate immediately	

4. Detect Characteristic AI Vocabulary and Phrase Patterns.

LLMs display lexical patterns. Common lexical patterns detected by NLP specialists in AI-generated text include: “delve into”, “in the realm of”, “it is crucial to note”, “a nuanced approach”, “as we navigate”, “tapestry of”, “let us explore” and “it is worth noting”. Apply an AI detection tool like GPTZero (gptzero.me), Copyleaks AI Detector or Originality.AI. GPTZero’s perplexity and burstiness scores should be viewed as indicative tools, not standardised threshold values.

POINT 4 – Characteristic AI Vocabulary & Detection Tools			
DOCUMENTED LLM LEXICAL SIGNATURES			
"delve into"	"in the realm of"	"It is crucial to note"	"a nuanced approach"
"as we navigate"	"tapestry of"	"let us explore"	"it is worth noting"
RECOMMENDED DETECTION TOOLS:			
GPTZero gptzero.me Perplexity + burstiness scores	Copyleaks copyleaks.com AI content probability	Originality.AI originality.ai Statistical LLM detection	

5. Verify Document Metadata and File Properties.

The information of Author, Created, Modified and Revision history can be found in File > Info > Properties within Microsoft Word. An example of document with claim of being a military intelligence brief yet its metadata reveals that it was only created 27 minutes prior to its distribution, created by User123, with Revision of ‘1’ and Application as python-docx/LibreOffice Writer.

POINT 5 — Document Metadata & File Properties Verification

File > Info > Properties (MS Word)		
Author	User123	⚠
Created	27 minutes before circulation	⚠
Last Modified	Same timestamp as creation	⚠
Revision Number	1 (never edited)	⚠
Application	python-docx / LibreOffice Writer	⚠

EXTRACTION TOOL

ExifTool

Extracts all embedded metadata from DOCX, PDF, image, and other file formats.

Command: `exiftool <filename>`

6. Assess Internal Temporal and Logical Consistency.

Create a timeline of the chronological sequence of events mentioned in the suspicious document. The AI-written document typically includes some kind of contradiction within it; for instance, mentioning an event that happened after the date of the document's supposed writing or mentioning different casualty numbers in different sections of the document or naming organisations that are both active and inactive.

POINT 6 — Internal Temporal & Logical Consistency

CONSTRUCT A DOCUMENT TIMELINE — Flag these contradictions:

Event Reference	⚠ References events AFTER document's supposed creation date
Casualty Figures	⚠ Different paragraph = different numbers for same incident
Organisation Status	⚠ Entity described as both 'active' and 'defunct' in same doc

Root cause: LLMs generate probabilistically — they do NOT perform logical consistency checks across paragraphs.

7. Cross-Reference Against Verified Institutional Sources.

Any document purporting to be an official government, military or intelligence communication must be cross-referenced against publicly accessible official sources: PIB releases (pib.gov.in), Ministry of External Affairs statements (mea.gov.in), official unit press communications and verified institutional websites. Forged documents typically replicate general document format but fail on specific institutional abbreviation usage, procedural language and authorised signatory patterns.

POINT 7 — Cross-Reference Against Verified Institutional Sources

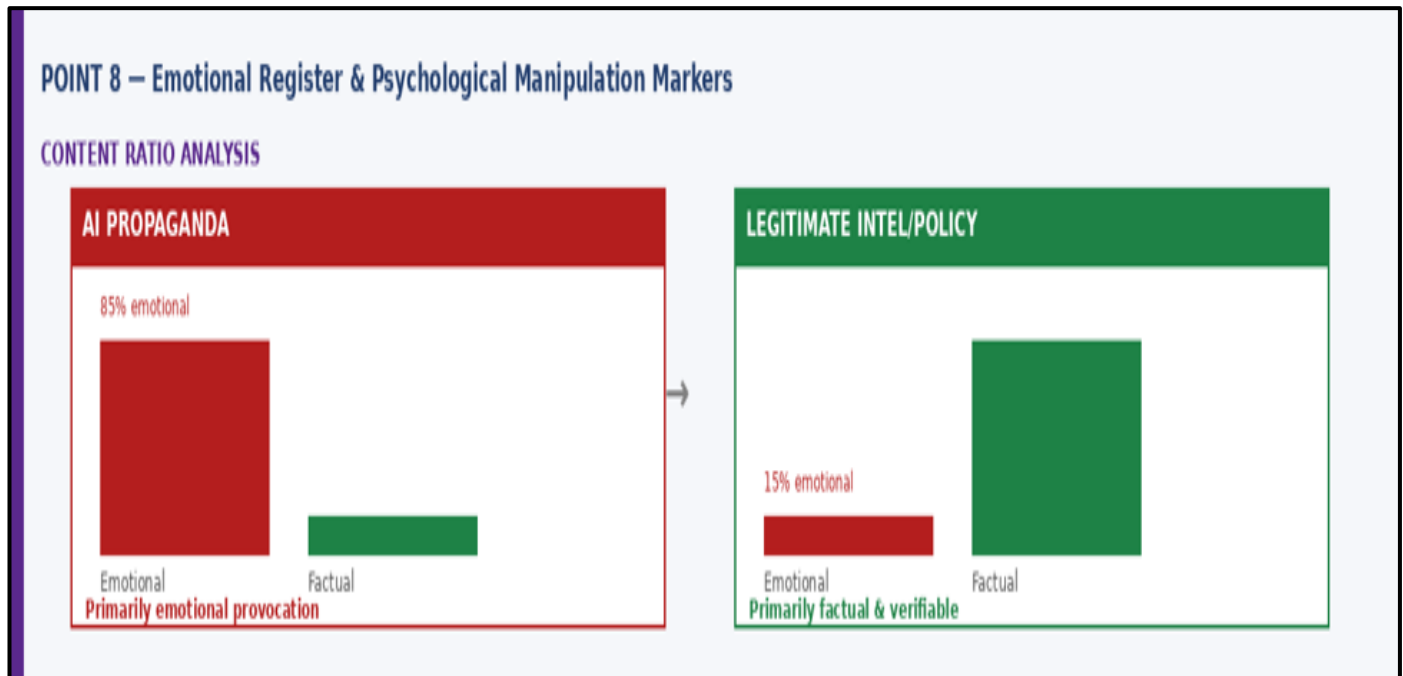
PRIMARY VERIFICATION SOURCES

PIB	pib.gov.in Press Information Bureau — official Govt of India releases	MEA	mea.gov.in Ministry of External Affairs statements
MOD	mod.gov.in Ministry of Defence — official defence communications	Army	indianarmy.nic.in HQ Indian Army official website & press releases

⚠ Forged documents replicate format but FAIL on: abbreviation usage • procedural language • authorised signatory patterns

8. Analyse Emotional Register and Psychological Manipulation Markers.

AI-generated propaganda text is typically calibrated to maximise emotional provocation. Examine the ratio of emotional content i.e. fear, humiliation, righteous anger, communal threat; to factual, verifiable information. Legitimate intelligence and policy documents are primarily factual; propaganda is primarily emotional. Systematic use of emotionally charged community-identity language without corresponding factual grounding is a reliable diagnostic.



9. Trace the Document's Dissemination Pathway.

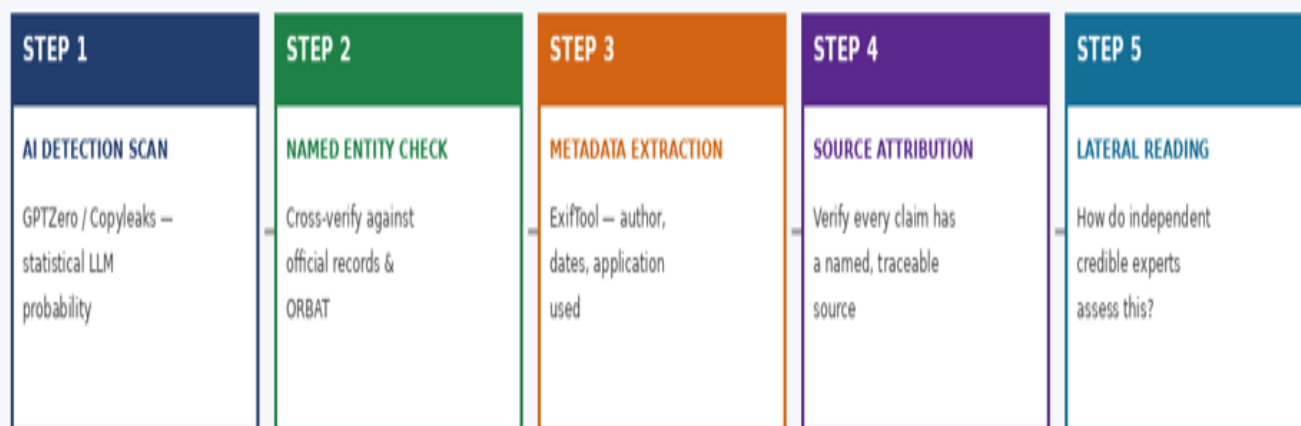
Investigate the chain of custody i.e. where did this document first appear? Anonymous Telegram channels, newly-created pastebins or accounts created within the previous week seeding content that spreads rapidly through informal networks are high-probability indicators of synthetic origin. Documents of genuine institutional origin have traceable provenance through established official communication channels, not anonymous shares.

POINT 9 – Trace the Document's Dissemination Pathway	
HIGH-RISK DISSEMINATION INDICATORS	LEGITIMATE PROVENANCE
☐ Anonymous Telegram channel No verifiable owner or admin	✓ Traceable to official .gov.in / .nic.in domain
☐ Newly-created pastebin Account age < 7 days	✓ Published through established press channels (PIB, ANI, PTI)
☐ Rapid viral spread Through informal networks only	✓ Consistent with prior communications from same source
☐ No institutional origin Cannot trace back to official domain	✓ Corroborated by independent credible outlets

10. Apply the Full OSINT Verification Stack Before Actioning.

Before treating any unsolicited document as operationally actionable, apply the five-step OSINT verification stack: firstly, AI detection tool scan via GPTZero or Copyleaks; secondly, Named entity cross-check against official records; thirdly, Metadata extraction via ExifTool; fourthly, Source attribution verification and lastly, Lateral reading i.e. examine how independent credible experts discuss the claims, not merely the document itself. A document failing two or more of these checks must be treated as potentially synthetic and escalated immediately.

POINT 10 – The 5-Step OSINT Verification Stack



⚠ Document FAILING 2 or more steps = treat as SYNTHETIC — escalate immediately through command chain.