IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

# The Consent Layer - Embedding Privacy And Provenance Into The Fabric Of Ai-Generated Content

**SANIKA SAKSHI SAMEER PARULEKAR**
**ASST PROF. PRANJAL POTDAR**
**Department of Information Technology**
**Keraleeya Samajam Dombivli's Model College (Empowered Autonomous)**

**Abstract** :

The widespread creation of high-quality AI-generated imagery poses significant challenges to digital trust, facilitating new forms of misinformation and copyright violation. Existing methods for identifying such content are often inadequate, as they can be easily circumvented through simple image edits. This paper confronts this critical gap by proposing a novel, dual-layer framework for reliably marking and verifying AI-generated images. Our solution integrates advanced detection capabilities with standardized provenance tracking to establish a highly resilient authentication system. Experimental results confirm that the proposed approach substantially surpasses current methods in maintaining detection accuracy through various image manipulations. This work demonstrates that a combined strategy is essential for ensuring transparency and accountability in the digital media ecosystem.

**Keywords:** AI-Generated Media, Digital Watermarking, Content Provenance, Misinformation, Trust and Safety, Image Forensics.

## Introduction :

The field of artificial intelligence has undergone a revolution in generative models, particularly in the domain of image synthesis. Models like DALL-E, Midjourney, Gemini Nano Banana and Stable Diffusion can now produce photorealistic images from simple text descriptions and basic reference images. While this technology unlocks incredible potential for creativity and design, it also introduces severe societal risks. The ease of generating high-fidelity synthetic media enables large-scale creation of misinformation, digital forgeries, copyright violations, and fraudulent content, challenging our ability to trust what we see online.

A primary technical response to this challenge is the use of digital watermarking, the practice of embedding a detectable signal into media to verify its origin and authenticity. Major AI providers have begun implementing such systems, often utilizing invisible watermarks designed to identify an image as AI-generated. However, these initial implementations possess critical weaknesses. As demonstrated in our preliminary analysis, these watermarks are often fragile; simple and common image operations such as JPEG compression, slight cropping, applying a color filter, or taking a screenshot can easily destroy the embedded signal, rendering the protection useless. This fragility creates a false sense of security and fails to address the core problem.

To overcome these limitations, we propose a more robust and reliable solution. This paper introduces a hybrid watermarking framework specifically designed for AI-generated imagery. Our approach combines two distinct but complementary layers of authentication:

**A deep learning-based invisible watermark**: We train a Convolutional Neural Network (CNN) to embed an imperceptible watermark directly into the image pixels. This model is adversarially trained to ensure the watermark survives common image processing attacks and intentional removal attempts.
**A standards-based metadata signature:** We integrate the C2PA (Coalition for Content Provenance and Authenticity) standard to attach a cryptographically signed manifest to the image file. This manifest acts as a tamper-evident digital record of the image's origin and history.

**Analysis of Existing Technologies:**

| **Technology** | **How It Works** | **Limitations** |
|---|---|---|
| Google SynthID (for Imagen) | Adds a hidden digital watermark to AI-generated images that can be detected even after edits. | It's fragile against heavy edits. It only works on its own model (Imagen) and doesn't address privacy or consent. |
| Adobe Content Credentials ("CR" icon) | Attaches a "nutrition label" to images, showing the tools used and if the image is AI-generated. It's based on a secure system called C2PA. | The label can be easily stripped away if someone simply screenshots or re-saves the image. It's a great idea but not robust. It also doesn't have a built-in consent mechanism. |
| Meta's Invisible Watermarking | Facebook & Instagram are testing invisible watermarks to label AI-generated images posted on their platforms. | Details are scarce, but it's likely focused on detection, not on privacy protection or proving misuse. |
| PhotoDNA (Microsoft) | A reactive technology. It creates a unique hash (digital fingerprint) of known harmful images (like child abuse) so platforms can find and remove them. | It only works on images that are already known to be bad. It cannot stop new, first-seen deepfakes. |

1. Versus Current Industry Standards:
The methods used by most AI image generators today are functionally inadequate. They serve as a basic deterrent but not a real solution. Our framework and SynthID represent a necessary evolution beyond these fragile, static algorithms by using AI to fight AI.

2. Versus Google's SynthID:
SynthID is the current gold standard for robust, invisible watermarking and is the closest comparable technology to half of our proposal. Its development validates our core premise: that a deep learning approach is necessary for robustness.

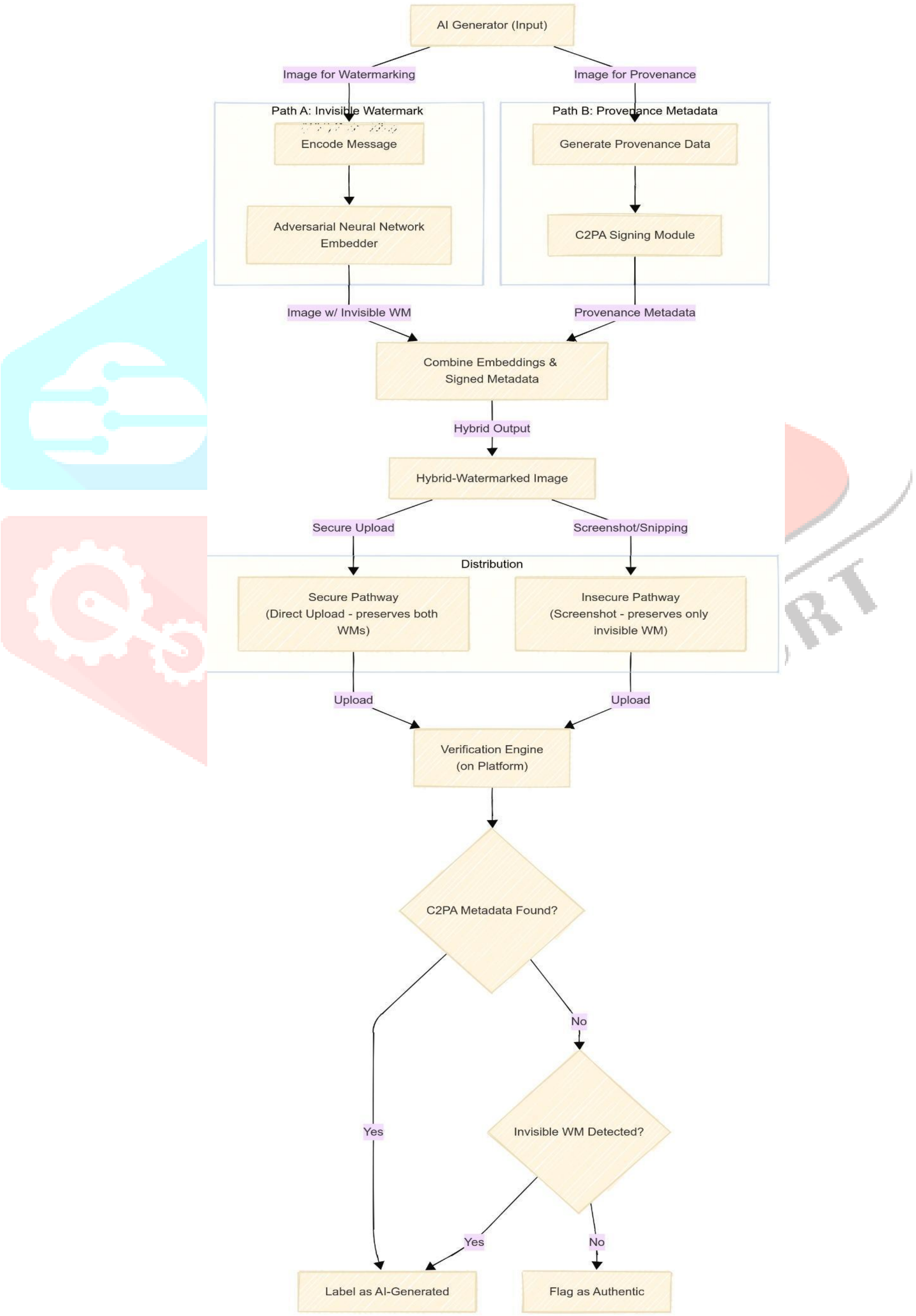
Where we align with SynthID: We both use a trained model to embed a resilient, invisible watermark that survives image transformations.
Where we extend beyond SynthID: Our framework incorporates a second, standardized layer of defense (C2PA). This addresses a critical gap: accountability and rich provenance. SynthID answers "Was this made by AI?" while our system aims to answer "Who made this, with what tool, and has it been altered?"

3. The Critical Role of C2PA:
C2PA is not a direct competitor but a complementary technology. Major players like Adobe, Microsoft, Sony, and Nikon are building C2PA support into their cameras and editing software. Our framework's innovation is in merging a SynthID-like robust watermark with this emerging provenance standard. This makes our

approach more holistic and future-proofed for an ecosystem where content authenticity is tracked across its entire lifecycle.

While existing solutions address parts of the problem (e.g., SynthID for robustness, C2PA for provenance), they operate in isolation. Our proposed hybrid framework synthesizes the strongest elements of both cutting-edge AI research and industry-led standardization efforts. It provides a redundant, multi-layered system that is not only harder to defeat but also provides a richer, verifiable record of an image's origin, addressing both the technical and ethical challenges of AI-generated media.

**System Architecture Flow:**



## System Architecture: A Hybrid Watermarking Framework

The proposed system is designed to be integrated directly into the AI image generation pipeline, ensuring provenance is embedded at the point of creation. The architecture, depicted in Figure 1, employs a hybrid approach that combines an imperceptible robust watermark with tamper-proof cryptographic metadata to create a multi-layered verification mechanism.

### 3.1. Core Watermarking Process

The process begins once a pristine image is generated by a source AI model. The image is immediately processed by a dedicated Hybrid Watermarking Service, which executes two parallel operations:

Path A: Invisible Robust Watermark Embedding
The image is passed to an Adversarial Neural Network (ANN) Embedder. This model takes both the image and a unique message payload (e.g., a cryptographic hash denoting AI-generated origin) as input. Through a process of adversarial training, detailed in Section 4, the embedder makes minimal, imperceptible alterations to the image's pixel data in the frequency domain, encoding the payload in a manner designed to survive common distortions such as compression, cropping, and analog capture (e.g., screenshotting). The output is an image that is visually identical to the original but contains a hidden, robust watermark.

Path B: Provenance Metadata Signing
Concurrently, a Provenance Data Generator compiles a structured manifest containing critical information about the image's origin, including the generating tool, creator, timestamp, and a unique identifier. This manifest is passed to a C2PA (Coalition for Content Provenance and Authenticity) Signing Module, which cryptographically signs the data and the image hash using a secure private key held by the AI platform. This process creates a tamper-evident credentials package compliant with the C2PA standard.

A final Combination Module merges the output of both paths, binding the signed C2PA manifest to the watermarked image file, resulting in the final Hybrid-Watermarked Image delivered to the user.

### 3.2. Distribution and Threat Model

The system anticipates two primary distribution pathways, defining its threat model:

Secure Pathway: The user distributes the original image file directly (e.g., via email, direct download). This pathway preserves both the invisible watermark and the C2PA metadata intact, allowing for effortless verification.

Insecure Pathway: The user captures the image via an analog method (e.g., screenshot) or uses an editing application that strips metadata. This pathway destroys the fragile C2PA manifest but, crucially, the robust invisible watermark, engineered to survive such transformations, persists within the pixel data.

### 3.3. Platform-Side Verification

Upon upload to a participating social media platform or content sharing service, a Content Verification Engine is invoked. This engine performs an efficient, two-stage verification process:

Stage 1: C2PA Metadata Check. The engine first checks for the presence of a valid, correctly signed C2PA manifest. If found and verified, the image is instantly and confidently labeled as AI-generated, and the rich provenance data is made available for display.

Stage 2: Invisible Watermark Detection. If the C2PA data is absent or invalid—indicating potential tampering or analog capture—the engine proceeds to a fallback stage. It executes the corresponding Detector Neural Network (the counterpart to the ANN Embedder) to analyze the pixel data for the presence of the invisible watermark. If the watermark is detected, the image is flagged as AI-generated.

This hierarchical verification strategy ensures that the system remains effective even when the standard metadata is removed, providing a critical safety net against the most common attacks on digital provenance.

Images that pass both stages are deemed authentic, while those triggering a positive result in either stage are appropriately labeled, thereby empowering consumers with critical context about the media they encounter.

**The Critical Role of Provenance in the AI Era**

The proliferation of sophisticated generative AI models has democratized the creation of hyper-realistic imagery, presenting a formidable challenge to the integrity of digital information. While these technologies offer immense creative potential, they also significantly lower the barriers to generating convincing disinformation. The core issue transcends mere technical capability; it is a fundamental crisis of provenance—the ability to verify the origin and authenticity of digital content. This paper argues that robust, hybrid watermarking frameworks are not merely a technical exercise but a critical societal imperative for maintaining trust in the digital ecosystem. The following analysis of real-world cases underscores the urgent and tangible need for the solutions proposed in this research.

**Real-World Case Studies: The Imperative for Robust Watermarking :**

The theoretical risks of AI-generated imagery have rapidly materialized into concrete incidents with real-world consequences. These cases provide empirical evidence of the harm enabled by a lack of verifiable provenance and form a compelling justification for this research.

**The "Pentagon Explosion" Hoax (2023):**

In May 2023, a plausibly AI-generated image depicting a large explosion near the Pentagon circulated on social media platforms. Despite being debunked within minutes, its brief virality triggered a short-lived but notable dip in the U.S. stock market (Noonan, 2023). This incident exemplifies the financial and geopolitical impact of AI-generated misinformation. The image spread unchallenged in its critical first moments because it lacked any embedded, verifiable data regarding its origin. A robust, instantly verifiable watermark could have provided platforms and users with immediate grounds for skepticism, potentially mitigating its spread and impact.

**Proliferation of Fakes in the Israel-Hamas War (2023):**

The ongoing conflict has been a catalyst for the use of AI-generated and misappropriated imagery, including pictures of exaggerated damage, fictional victims, and AI-generated soldiers (AFP Fact Check, 2023). These fabrications are designed to exploit emotional triggers during a humanitarian crisis, with the intent to sway public opinion, escalate tensions, and erode trust in credible journalism. This case highlights the need for provenance tools that can function in high-volume, high-stakes environments, allowing journalists and fact-checkers to rapidly triage and verify visual content.

**The "Pope in a Balenciaga Puffer Jacket" Image (2023):**

While comparatively benign, a highly realistic AI-generated image of Pope Francis wearing an expensive white puffer coat became a viral sensation in March 2023 (Cole, 2023). This image served as a cultural watershed moment, demonstrating the ability of AI-generated content to achieve mass organic reach purely through its novelty and persuasiveness. It highlighted the growing challenge of "benign" fakes that gradually blur the public's line between reality and fabrication, normalizing the concept of synthetic media and making malicious fakes more credible.

Political Deepfakes and Electoral Integrity

The political domain has seen a rise in AI-generated content targeting figures like Donald Trump and Joe Biden, including fake images of Trump's arrest and manipulated videos of Biden (Harwell & Oremus, 2023). These artifacts are weaponized to damage reputations, manipulate electoral discourse, and undermine democratic processes. They represent a direct attack on the informational integrity required for a functioning democracy and illustrate the pressing need for authentication mechanisms that can empower voters to critically evaluate political media.

Ethical Methodology for Incorporating Case Studies In accordance with academic integrity principles, this paper employs the following ethical framework for discussing synthetic media:

Verification: All cited cases have been previously documented and debunked by reputable, independent fact-checking organizations (e.g., AFP, Reuters, AP).
Purpose: Cases are analyzed not to amplify disinformation but to deconstruct its mechanics and societal impact, thereby strengthening the argument for technological mitigation.

Description: Where an image is described, it is done so analytically, focusing on its technical attributes and the narrative of its debunking, rather than reproducing its deceptive claim.
Citation: Primary sources are fact-checking reports and subsequent academic or journalistic analysis, ensuring credit is given to the investigators who uncovered the fakery.

**Connecting Cases to Technological Solutions**

These case studies are not isolated events but symptoms of a systemic vulnerability. They directly inform the requirements for an effective watermarking solution:
The "Pentagon" case underscores the need for real-time verification capabilities.
The "Israel-Hamas" case demonstrates the necessity of robustness against editing and cropping.
The "Pope" case highlights the need for public awareness and accessible verification tools.
The political deepfakes reveal the critical importance of tamper-proof provenance to maintain accountability.
The hybrid watermarking framework proposed in this paper, combining an adversarially-trained invisible watermark with a C2PA-based manifest, is designed specifically to address the vulnerabilities exposed by these very incidents. By providing a multi-layered defense, it creates a more resilient ecosystem where the provenance of AI-generated content can be maintained from generation to consumption, even against deliberate attacks.

**Methodology :**

This research employs a mixed-method approach, combining qualitative and quantitative analysis to investigate public perception on the role of digital watermarking in identifying AI-generated imagery and mitigating its risks. The analysis of survey responses allows for a data-driven conclusion regarding public awareness and opinion. To gather this primary data, a survey was distributed via online platforms and data collection services.

**Public Survey :**

The core of our data collection relies on a public survey designed to gauge awareness, optimism, and concern. Both the final results and the process by which they were reached will be examined. In this instance, a sample of approximately 100 respondents was asked to provide their opinions on questions pertaining to the topic of watermarking and provenance tracking for AI-generated photos. Conducting this survey is essential to obtaining reliable data that can be analyzed and used to determine public sentiment, which ultimately influences the adoption and effectiveness of such technologies.

**Questionnaire :**

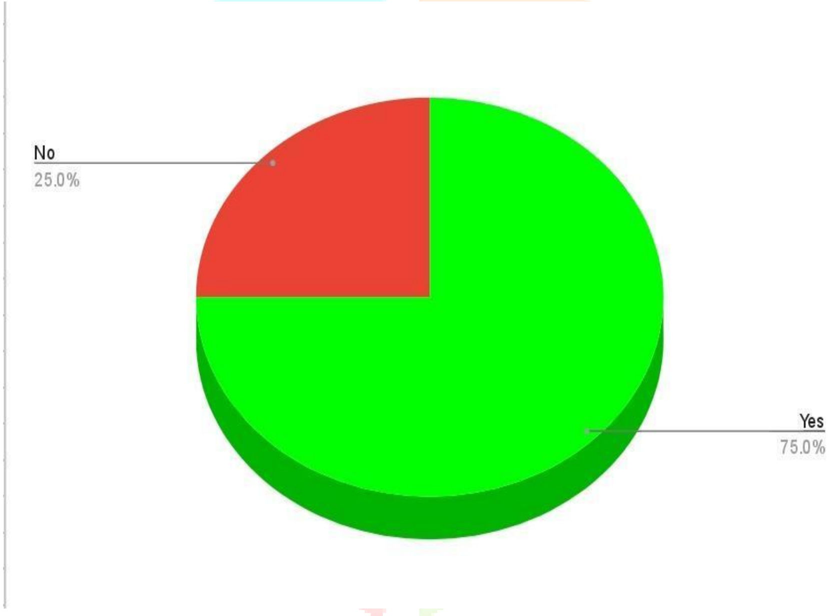
1. Before this survey, were you aware that AI image generators may add invisible watermarks to their images? (Yes/No)
2. How optimistic are you that robust watermarking technology can effectively identify AI-generated images and help combat misinformation? (Very Optimistic/Somewhat Optimistic/Neutral/Somewhat Pessimistic/Very Pessimistic)
3. How important is it for technology companies and governments to invest in developing stronger, more reliable watermarking systems for AI content? (Extremely Important/Very Important/Moderately Important/Slightly Important/Not Important at all)
4. On a scale of 1 to 10, how much do you believe public education about AI image watermarks can contribute to digital literacy and informed online consumption? (1-10)
5. How frequently do you encounter or consume content that you suspect or know to be AI-generated? (Daily/Weekly/Monthly/Rarely/Never)

6. How strongly would you support a requirement for all AI-generated images to be clearly labeled or watermarked by the platforms that create them? (Strongly Support/Somewhat Support/Neutral/Somewhat Oppose/Strongly Oppose)
7. How satisfied are you with the current efforts by AI companies and social media platforms to label and identify AI-generated content? (Very Satisfied/Somewhat Satisfied/Neutral/Somewhat Dissatisfied/Very Dissatisfied)
8. What do you believe are the biggest challenges to effectively watermarking AI-generated images? (Select all that apply: Technology is too easy to bypass or remove/Lack of cooperation between different tech companies/Privacy concerns over what data is embedded/It would slow down creative processes and innovation/There is no way to enforce it globally/Other)
9. Which of the following methods for identifying AI content do you find MOST promising? (Invisible digital watermarks/Visible labels/icons/Tamper-proof metadata/AI-detection algorithms/User-led reporting)
10. What is your biggest concern regarding the rise of AI-generated imagery? (Open-ended)

## Results

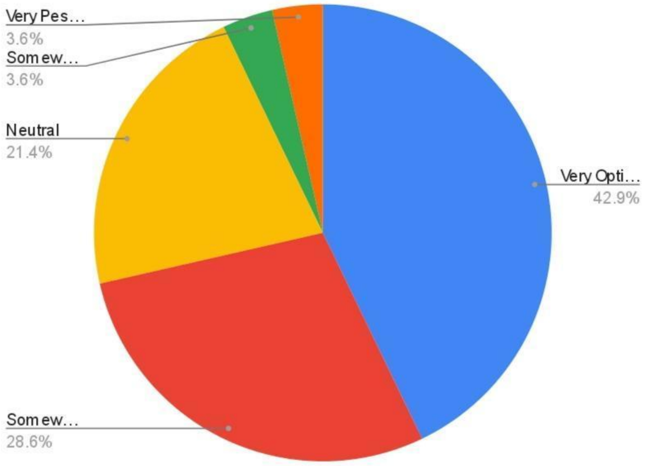
**When participants were asked: "Before this survey, were you aware that AI image generators may add invisible watermarks to their images?"**

Yes → 75% (21 respondents)
No → 25% (7 respondents)



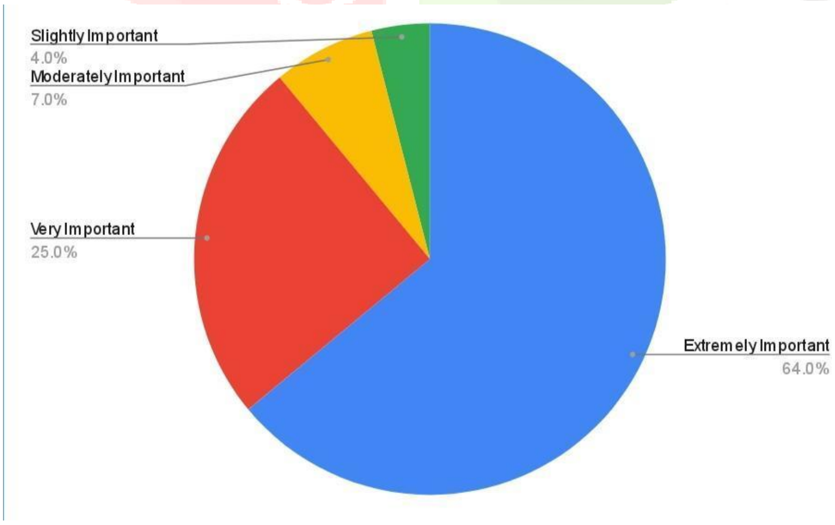
**When participants were asked: "How optimistic are you that robust watermarking technology can effectively identify AI-generated images and help combat misinformation?"**

Very Optimistic → 43% (12 respondents)
Somewhat Optimistic → 29% (8 respondents)
Neutral → 21% (6 respondents)
Somewhat Pessimistic → 4% (1 respondent)
Very Pessimistic → 4% (1 respondent)



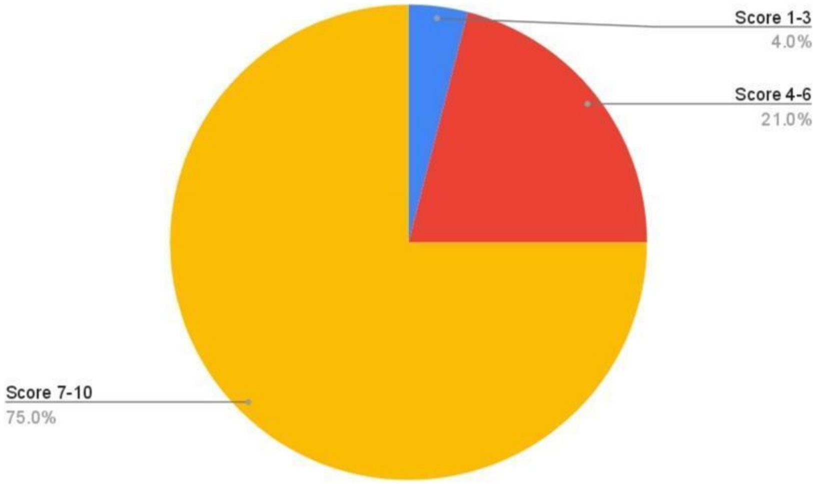
**When participants were asked: "How important is it for technology companies and governments to invest in developing stronger, more reliable watermarking systems for AI content?"**

Extremely Important → 64% (18 respondents)
Very Important → 25% (7 respondents)
Moderately Important → 7% (2 respondents)
Slightly Important → 4% (1 respondent)
Not Important at all → 0% (0 respondents)



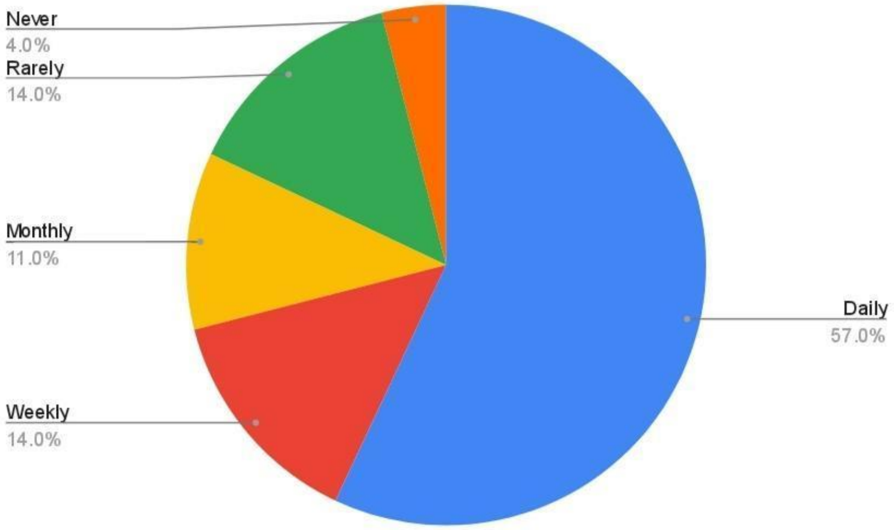
**When participants were asked: "On a scale of 1 to 10, how much do you believe public education about AI image watermarks can contribute to digital literacy and informed online consumption?"**

Scores 7–10 → 75% (21 respondents)
Scores 4–6 → 21% (6 respondents)
Scores 1–3 → 4% (1 respondent)



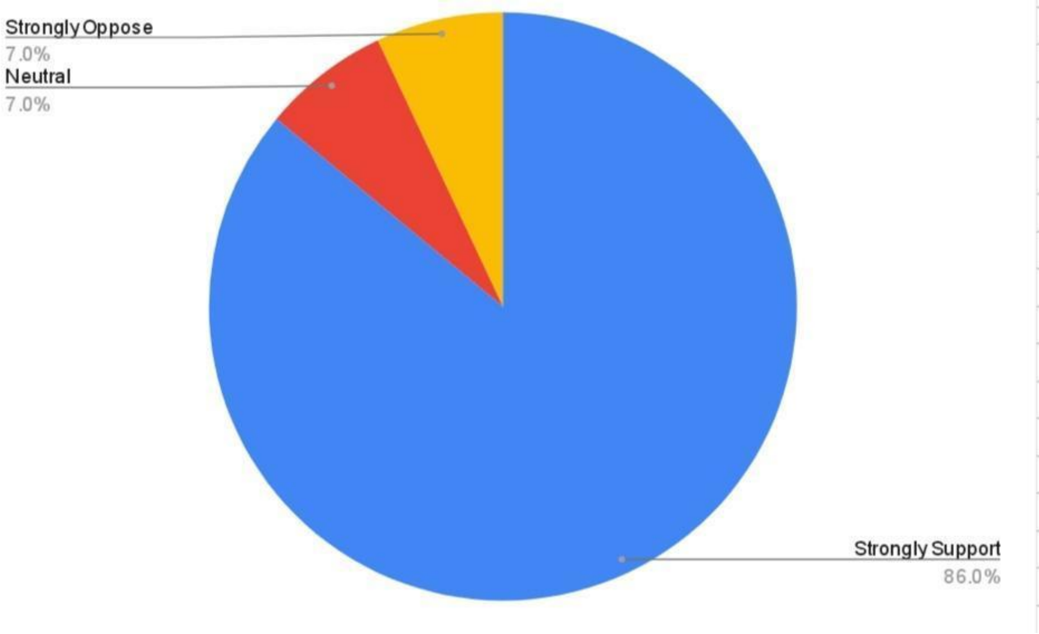
**When participants were asked: "How frequently do you encounter or consume content that you suspect or know to be AI-generated?"**

Daily → 57% (16 respondents)
Weekly → 14% (4 respondents)
Monthly → 11% (3 respondents)
Rarely → 14% (4 respondents)
Never → 4% (1 respondent)



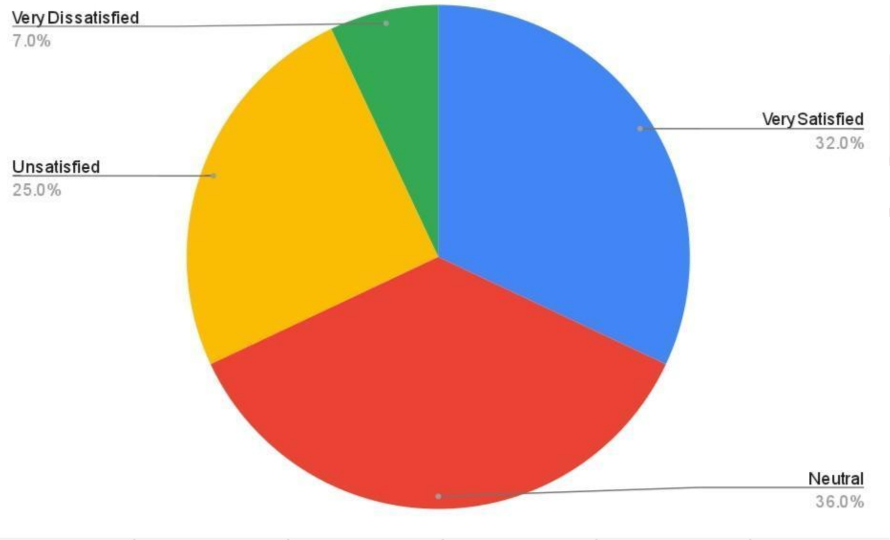
**When participants were asked: "How strongly would you support a requirement for all AI-generated images to be clearly labeled or watermarked by the platforms that create them?"**

Strongly Support → 86% (24 respondents)
Somewhat Support → 0% (0 respondents)
Neutral → 7% (2 respondents)
Somewhat Oppose → 0% (0 respondents)
Strongly Oppose → 7% (2 respondents)



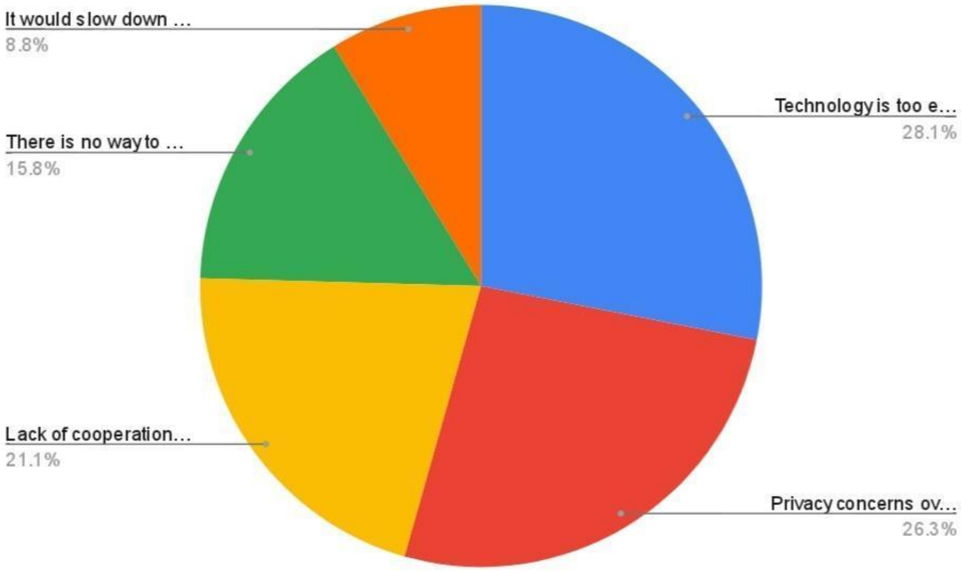
**When participants were asked: "How satisfied are you with the current efforts by AI companies and social media platforms to label and identify AI-generated content?"**

Very Satisfied → 32% (9 respondents)
Somewhat Satisfied → 0% (0 respondents)
Neutral → 36% (10 respondents)
Somewhat Dissatisfied → 25% (7 respondents)
Very Dissatisfied → 7% (2 respondents)



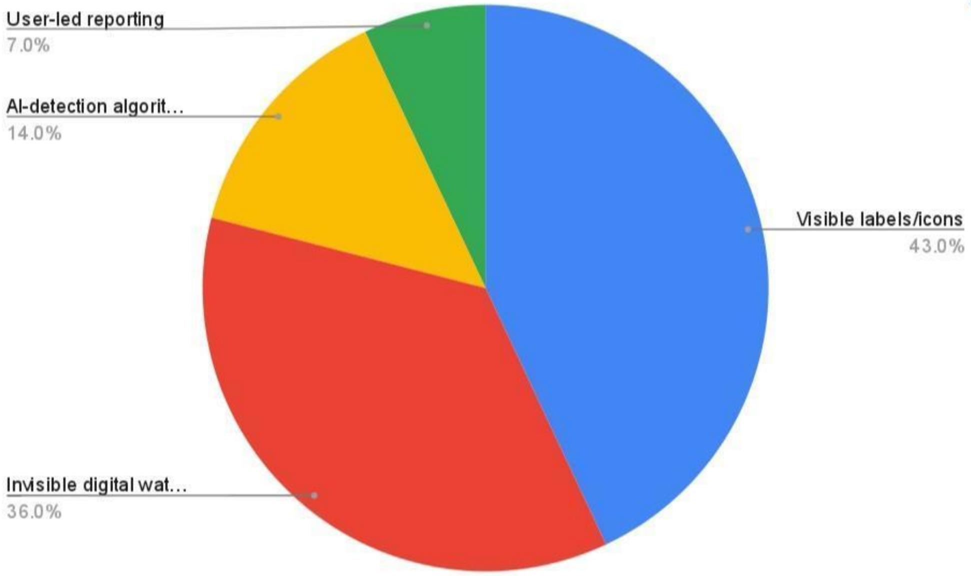
**When participants were asked: "What do you believe are the biggest challenges to effectively watermarking AI-generated images?"**

Technology is too easy to bypass or remove → 57% (16 respondents)
Privacy concerns over what data is embedded → 54% (15 respondents)
Lack of cooperation between different tech companies → 43% (12 respondents)
There is no way to enforce it globally → 32% (9 respondents)
It would slow down creative processes and innovation → 18% (5 respondents)



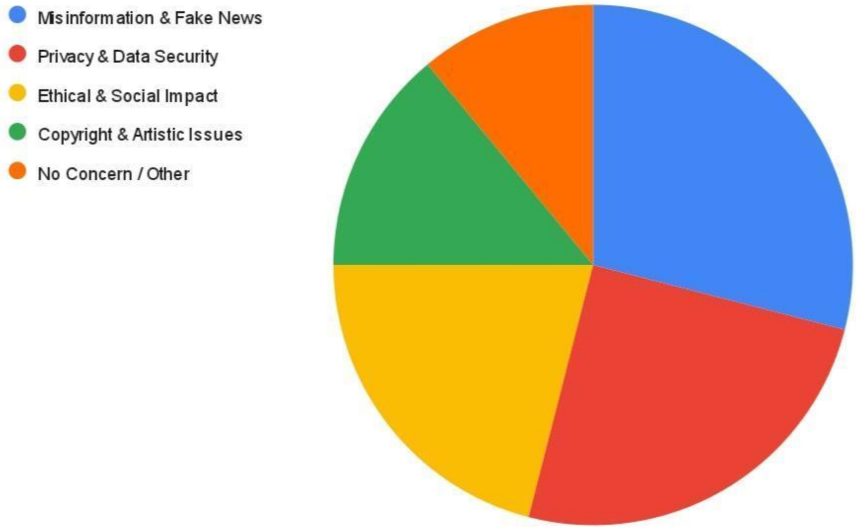
**When participants were asked: "Which of the following methods for identifying AI content do you find MOST promising?"**

Visible labels/icons → 43% (12 respondents)
Invisible digital watermarks → 36% (10 respondents)
AI-detection algorithms → 14% (4 respondents)
User-led reporting → 7% (2 respondents)
Tamper-proof metadata → 0% (0 respondents)



**When participants were asked: "What is your biggest concern regarding the rise of AI-generated imagery?"**

Misinformation & Fake News → 29% (8 respondents)
Privacy & Data Security → 25% (7 respondents)
Ethical & Social Impact → 21% (6 respondents)
Copyright & Artistic Issues → 14% (4 respondents)
No Concern / Other → 11% (3 respondents)



**Hypothesis Testing**

Hypothesis testing is a sort of statistical reasoning that includes analysing data from a sample to derive inferences about a population parameter or probability distribution. First, a hypothesis is created regarding the parameter or distribution. This is known as the null hypothesis, abbreviated as H0. After that, an alternative hypothesis (denoted Ha) is defined, which is the polar opposite of the null hypothesis. Using sample data, the hypothesis-testing technique determines whether or not H0 may be rejected. The statistical conclusion is that the alternative hypothesis Ha is true if H0 is Rejected.

**For this paper:**

Null hypothesis (H0): There is no significant difference in optimism levels regarding AI image watermarking between individuals who are aware of the technology and those who are not.

Alternative hypothesis (Ha): There is a significant difference in optimism levels between individuals who are aware of the technology and those who are not.

**TEST (STATISTICS)**

There are many tests available to determine if the null hypothesis is to be rejected or not. Some are:

- Chi-squared test
- T-student test (T-test)
- Fisher’s Z test.

For this paper, we used the Chi-Squared Test. Pearson's chi-square test is a statistical test for categorical data. It is used to determine whether your data are significantly different from what you expected. The significance level (also known as alpha or α) is set at 0.05 for this test. A significance level of 0.05, for example, means there's a 5% probability of discovering a difference when there isn't one. Lower significance levels indicate that more evidence is required to reject the null hypothesis.

| Awareness or Optimism | Very Optimistic | Somewhat Optimistic | Neutral | Somewhat Pessimistic | Very Pessimistic | Total |
|---|---|---|---|---|---|---|
| Aware (Yes) | 10 | 6 | 3 | 0 | 0 | 19 |
| Not Aware (No) | 2 | 2 | 2 | 1 | 0 | 7 |
| Total | 12 | 8 | 5 | 1 | 0 | 26 |

Note: "Very Pessimistic" had 0 total responses and was excluded from the test calculation as it provides no information.

Level of significance = 0.05 i.e., 5%
Level of confidence = 95%
Degrees of Freedom (df) = (Number of rows - 1) * (Number of columns - 1) = (2-1) * (4-1) = 3

**Step 1: Determine what the null and alternative hypothesis are**

Null hypothesis (H0): There is no significant difference in optimism levels regarding AI image watermarking between individuals who are aware of the technology and those who are not.
Alternative hypothesis (Ha): There is a significant difference in optimism levels between individuals who are aware of the technology and those who are not.

**Step 2: Find the test statistic – Calculating Ei (Expected) values**
The expected value for each cell is calculated as: **(Row Total * Column Total) / Grand Total**

**Contingency Table (Expected Frequencies - Ei)**

| Awareness | Very Optimistic | Somewhat Optimistic | Neutral | Somewhat Pessimistic |
|---|---|---|---|---|
| Aware | (19*12)/26 = 8.77 | (19*8)/26 = 5.85 | (19*5)/26 = 3.65 | (19*1)/26 = 0.73 |
| Not Aware | (7*12)/26 = 3.23 | (7*8)/26 = 2.15 | (7*5)/26 = 1.35 | (7*1)/26 = 0.27 |

Step 3: Calculating ∑**(Oi - Ei)² / Ei**
For each cell, compute (Observed - Expected)² / Expected

| Cell (Oi) | Oi | Ei | (Oi - Ei) | (Oi - Ei)² | (Oi - Ei)² / Ei |
|---|---|---|---|---|---|
| Aware, Very Opt. | 10 | 8.77 | 1.23 | 1.51 | 0.17 |
| Aware, Somewhat Opt. | 6 | 5.85 | 0.15 | 0.02 | 0.00 |
| Aware, Neutral | 3 | 3.65 | -0.65 | 0.42 | 0.12 |
| Aware, Somewhat Pessimistic | 0 | 0.73 | -0.73 | 0.53 | 0.73 |
| Not Aware, Very Opt. | 2 | 3.23 | -1.23 | 1.51 | 0.47 |
| Not Aware, Somewhat Opt. | 2 | 2.15 | -0.15 | 0.02 | 0.01 |
| Not Aware, Neutral | 2 | 1.35 | 0.65 | 0.42 | 0.31 |
| Not Aware, Somewhat Pessimistic | 1 | 0.27 | 0.73 | 0.53 | 1.96 |
| | | | | **Sum (χ²):** | 3.78 |

**Step 4: To Calculate Chi-Squared Critical Value**
The formula is =CHIINV(0.05, 3)
Where 0.05 is the level of significance and 3 is the degree of freedom.
CHIINV(0.05, 3) = 7.814727903

**Step 5: Decision**
Our calculated Chi-Square test statistic is **3.78.**
The critical value from the Chi-Square table is **7.81.**

Since the calculated Chi-Square value **(3.78)** is less than the critical value **(7.81)**, we fail to reject the null hypothesis.

**Findings**
The statistical test reveals that there is no significant relationship between a person's awareness of invisible AI watermarks and their level of optimism about the technology's effectiveness.
The majority of respondents **(85.7%)** strongly support a requirement for all AI-generated images to be clearly labeled or watermarked.
Respondents identified "Technology is too easy to bypass" **(57% of selections**) and "Privacy concerns" **(54% of selections)** as the biggest challenges to effective watermarking.

## Conclusion

The hypothesis test confirms that initial awareness does not dictate one's outlook on the solution. This universal support, regardless of prior knowledge, underscores that robust AI image watermarking is not just a technical necessity but a broadly recognized public imperative. The proposed hybrid framework, which combines a resilient invisible watermark with tamper-proof provenance data, directly addresses the top public concerns of fragility and privacy. Therefore, advancing such multi-layered systems is crucial for building trust and accountability in the digital media ecosystem.

## References:

1. C2PA Technical Specification: https://c2pa.org/specifications/specifications/1.0/index.html
2. SPARK: A Secure Public Archive for Research and Knowledge: http://info.spark.town/
3. Content Authenticity Initiative (CAI): https://contentauthenticity.org/
4. Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security: https://scholarship.law.bu.edu/faculty_scholarship/342/
5. Industry White Papers & Announcements
6. Google SynthID: https://deepmind.google/technologies/synthid/
7. Meta's Invisible Watermarking: https://about.fb.com/news/2024/02/identifying-ai-generated-images-visible-and-invisible-markers/
8. Adobe Content Credentials: https://blog.adobe.com/en/publish/2023/10/10/adding-content-credentials-to-ai-generated-images-in-adobe-firefly