Journal of International DBA Studies (JIDS)
Vol. 2 No. 001 (2026) https://eaj.ebujournals.lu/index.php/JIDS/index

A Dual Strategy for Digital Market Integrity: Content Credentials and Consumer Trust

by
Janak Makwana *

Abstract

The rapid advancement of generative AI has raised concerns about visual trust in digital commerce, as highly realistic synthetic images make it challenging to distinguish between authentic and artificial content. This article examines implications for e-commerce and online food delivery, using India's fast-growing digital marketplace as context and extending to global policy trends. Grounded in signaling theory, it argues that AI-generated visuals act as low-cost, deceptive signals in information-asymmetric environments, eroding consumer confidence and inflating post-purchase dissatisfaction. Detection-only strategies fall short due to adversarial adaptation and model drift. It proposes the Verifiable Authenticity Framework: a phased playbook that (1) embeds Content Credentials via the C2PA standard across the media supply chain; and (2) ensures consumer-facing transparency at purchase, with optional cryptographic notarization for high-value categories. This enables brands to send costly, auditable signals of authenticity, restoring trust in image-driven markets. An experimental design is outlined to assess its impact on return rates, ratings, and repurchase intent. By embedding provenance and transparency, firms can reduce operational risk, align with emerging cross-jurisdictional mandates, and build sustainable competitive advantage in the global AI-driven economy.

Keywords: Content Credentials; Consumer Trust; Signaling Theory; Market Governance; International Policy; E-Commerce Operations.

*Janak Makwana is a DBA Researcher specializing in Generative AI, digital trust, and scalable frameworks to enhance transparency and integrity in AI-driven commerce. He is a technology leader with over two decades of experience across global technology and financial services organizations.

Introduction: A crisis of visual trust

Digital commerce is an increasingly image-led environment. Product photos and short videos often serve as the decisive signal of quality at the point of choice. As generative AI lowers the cost of photorealistic production, that "polish" can no longer be treated as a trustworthy proxy for underlying product quality. The result is a visual trust gap: consumers face information asymmetry at the point of purchase, while brands confront higher return rates, ratings volatility, and reputational exposure.

A vivid example of this crisis can be seen in the online food delivery space. Consider the case of Hyderabadi Biryani-a dish known for its rich aroma, layered textures, and regional authenticity. On many platforms, the listing image is a hyper-realistic, AI-enhanced visual that showcases perfectly arranged rice grains, vibrant garnishes, and steam effects. Yet, what arrives at the customer's doorstep may be a soggy, poorly packed version that bears little resemblance to the promise. This expectation-reality gap undermines consumer trust and fuels dissatisfaction.

Figure 1: A crisis of visual trust

AI-Generated **Authentic / Actual**

India offers a salient initial context. Multiple forecasts place the country's e-commerce market on a steep trajectory toward USD 300 billion+ by FY2030, propelled by smartphone penetration, low-cost data, and UPI adoption. For example, the India Brand Equity Foundation (IBEF) projects growth from approximately USD 125 billion in FY2024 to USD 345 billion by FY2030 (India Brand Equity Foundation, 2025). As volume scales, small percentage shifts in returns or in post-purchase satisfaction compound into material operational and P&L impacts. In the United States, the National Retail Federation (NRF) estimates total retail returns of USD 890 billion in 2024, with retailers expecting 16.9% of annual sales to be returned, illustrating how returns have become a structural feature of digital commerce economics (National Retail Federation, 2024).

At the same time, policy expectations are hardening. The EU AI Act (Regulation (EU) 2024/1689) has entered into force with staggered obligations, including transparency requirements for synthetic media and governance for general-purpose AI. Milestones began in February 2025, with additional provisions applying from August 2025 and further phases through 2026. Parliamentary materials reinforce that providers of AI systems must mark

outputs in a machine-readable format, so that they are detectable as artificially generated or manipulated (European Parliament, 2025). Regulators outside the EU are also active: the U.S. Federal Trade Commission (FTC) has warned that tools used to deceive, including deepfakes and voice clones, can trigger Section 5 enforcement on unfair or deceptive practices, an enforcement theme continued through 2024–2025 sweeps (Davis Polk, 2024). The UK Advertising Standards Authority (ASA) and the Committee of Advertising Practice (CAP) have likewise clarified that existing advertising rules already cover misleading or harmful AI-generated imagery and are ramping up proactive monitoring (Advertising Standards Authority, 2025).

Against this backdrop, detection-only controls are necessary but insufficient. As models improve and open-source ecosystems diffuse capabilities, adversarial adaptation erodes detector durability. A complementary path is to move from "trust me" to "see for yourself": make provenance and edit history verifiable, and surface that verification to consumers at the point of purchase. The Coalition for C2PA specification underpins Content Credentials cryptographically verifiable manifests that bind capture context and edit to the asset. The v2.1 technical specification focuses on reliability and security; implementations increasingly pair with watermarking to anchor provenance signals (Coalition for Content Provenance and Authenticity, 2025a, 2025b). Consumer-facing explainers and verification tools are now accessible, normalizing the "pin" metaphor (badge → modal → full history) (Adobe Inc., 2025).

This article's contribution is therefore managerial and methodological for an international business audience. It presents a dual strategy-C2PA-first provenance and consumer-visible transparency-organized as a phased implementation playbook; it also provides an experimental design that connects provenance deployment to key performance indicators such as returns and repurchase intent. Importantly, this article proposes a framework and outlines a testable evaluation design; it does not report executed experimental results.

Literature Review and Theoretical Lens

Signaling Theory and the Collapse of "Polish" as a Credible Cue

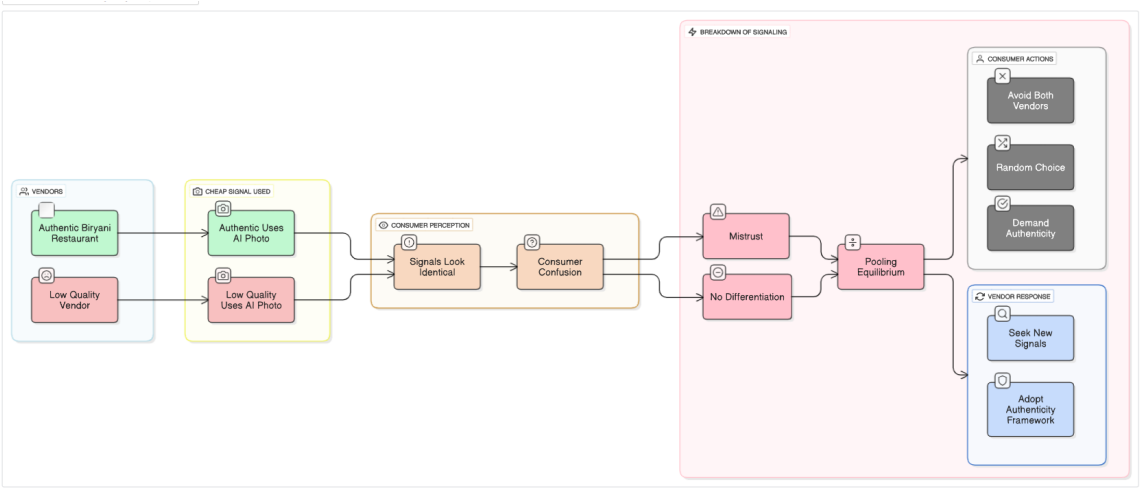
Markets with information asymmetry rely on observable signals that map credibly to unobservable quality. In classic signaling theory, costly and hard-to-imitate signals enable high-quality sellers to separate from low-quality rivals, restoring efficient market sorting (Spence, 1973). Historically, studio-grade visuals functioned as such signals in digital commerce: achieving high production values demanded time, expertise, and equipment, so "polish" conveyed professionalism and implicitly, quality.

However, the rise of generative artificial intelligence (AI) has dramatically reduced the cost of producing photorealistic imagery. This shift undermines the credibility of visual polish as a signal of quality. Sellers, regardless of actual product standards, can now deploy

high-quality visuals at minimal cost, leading to a pooling equilibrium whereby consumers struggle to distinguish between honest and deceptive sellers. Images become “cheap talk” rather than costly signals (Spence, 1973; Tolosana et al., 2020).

In such pooling settings, post-purchase outcomes deteriorate, return rates increase, ratings become more volatile, and electronic word of mouth (e-WOM) polarizes. Large-scale retail diagnostics in the United States illustrate the economic significance of this dynamic; retailers reported USD 890 billion in returns in 2024 and estimated that 16.9% of annual sales would be returned, a structural drag on margin for online categories that lean heavily on images (National Retail Federation, 2024). The managerial problem, therefore, is not merely one of content production. Rather, it is one of restoring credible differentiation at the Product Detail Page (PDP), where expectations are set, and conversion decisions are made (National Retail Federation, 2024; India Brand Equity Foundation, 2025). To address this issue, the Verifiable Authenticity Framework introduces costly, auditable signals such as cryptographically verifiable Content Credentials that deceptive sellers cannot easily mimic. This restores a separating equilibrium, where high-quality sellers can credibly signal integrity and consumers regain trust.

Figure 2: The Shift from a Separating to a Pooling Equilibrium and the Path to Restoration



Detection-Only Approaches: Necessary but Insufficient

A rich stream of research seeks to detect synthetic or manipulated media via statistical artifacts, forensic cues, or learned discriminators. However, detector performance tends to degrade as generative models evolve, attackers adapt to deployed defences, and class imbalance and dataset drift undermine calibration in production environments (Tolosana et al., 2020; Ghosh, Bellinger, Corizzo, Branco, Krawczyk, and Japkowicz, 2024). Given the rapid diffusion of model capabilities through open-source ecosystems and the speed of post-processing innovations, firms that rely solely upon

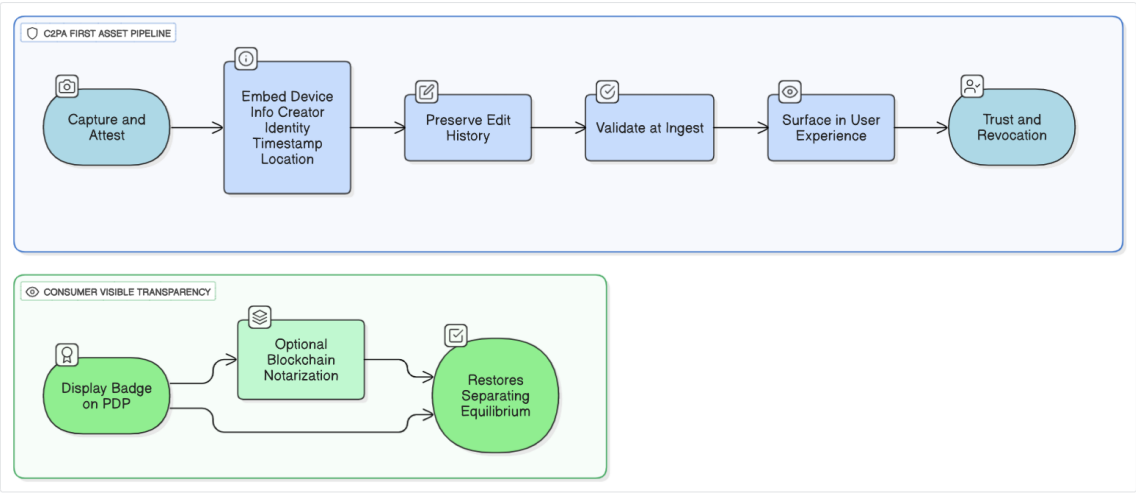
detection face a moving target, wherein adversarial adaptation and model drift can undermine reliability, and thus increase operational costs (Tolosana et al., 2020; Ghosh et al., 2024).The implication for practice is clear: detection is necessary, but it should be complemented by mechanisms that make honesty cheaper and verifiable, and deception harder and riskier.

Provenance as Infrastructure: C2PA and Content Credentials

To overcome these limitations, a complementary approach centers on provenance-cryptographically verifiable records that can attest to who captured content, when, and what edits occurred along the way. The C2PA specification defines embeddable manifests that travel with the asset across the supply chain and can be validated at ingest or render time (Coalition for Content Provenance and Authenticity, 2025a). In practice, these manifests underpin Content Credentials: a consumer-legible badge and modal that summarizes origin and edit history in plain language, with the option to explore a full, signed history (Adobe Inc., 2025; Content Credentials, 2025). Recent ecosystem updates highlight C2PA v2.1 security and reliability enhancements and the growing practice of pairing manifests with digital watermarking to strengthen the link between the visual asset and its provenance data across recompression and resharing (Coalition for Content Provenance and Authenticity, 2025a; Digimarc, 2024).

From a signaling theory perspective, provenance and Content Credentials reintroduce cost and auditability. To display a valid badge, brands must invest in authentic capture, preserve editing histories, maintain trust lists of acceptable signers, and pass validation at ingest and render. Opportunistic sellers that previously free-rode on synthetic polish face higher frictions or failed validation, while honest brands emit a public, verifiable signal of integrity at the PDP (Coalition for Content Provenance and Authenticity, 2025a; Adobe Inc., 2025).

Figure 3 illustrates the two pillars of the Verifiable Authenticity Framework: a C2PA-first asset pipeline and consumer-visible transparency on the PDP.



This visual shows how provenance and Content Credentials introduce cost and auditability, making it harder for opportunistic sellers to fake authenticity, while enabling honest brands to stand out through verifiable trust.

The Technology Landscape

The following table summarizes three strategic approaches to mitigating AI-generated deception in digital commerce. It compares reactive detection, provenance via the C2PA standard, and optional on-chain notarization, highlighting their operational mechanisms, strengths, and limitations.

Table 1: Strategic Comparison for Addressing AI-Generated Deception

Strategy	How It Works	Pros	Cons
Reactive Detection (AI Detectors)	Analyses images for statistical artifacts and inconsistencies indicative of AI generation.	Can be applied to existing content. Rapid initial implementation	Perpetual arms race with advancing AI. High false-positive/negative rates. Does not prevent deception, only identifies it.
Provenance (C2PA Standard)	Cryptographically binds verifiable metadata (origin, edit history) to content at the source.	Provides a proactive, permanent record. Open international standard. Allows for consumer-facing verification. Shifts burden from detection to proof.	Requires adoption across the content supply chain. New workflows and tooling required.
On-Chain Notarization	Anchors a cryptographic hash of the content provenance to a public blockchain.	Provides immutable, tamper-proof timestamping. Enhances trust for high-value items. Enables secondary market verification.	Higher complexity and cost. UX challenges for average consumers. Overkill for most consumer goods.

Conceptual model and study aim (expressed in prose)

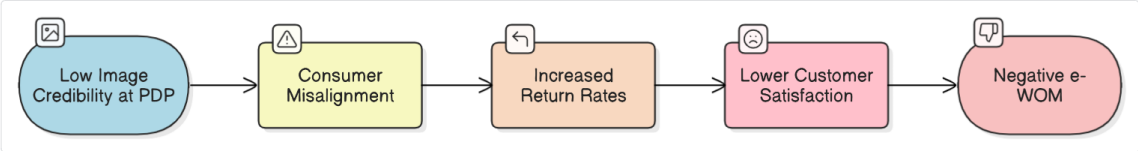
Taken together, the literature suggests that embedding provenance and surfacing verification on PDPs can establish separating equilibria by turning visuals into costly, credible, and auditable signals. In this article, we examine whether displaying Content Credentials badges on PDPs increases perceived authenticity and trust, reduces return rates, shifts ratings distributions upward, lowers image-related complaint rates, and improves repurchase behavior within 60–90 days. We also explore whether these effects vary by category (e.g., apparel/beauty/food, where expectation-setting via imagery is especially sensitive) and how policy and governance contexts interact with brand adoption

decisions. Consistent with signaling theory, we expect that credentialed PDPs will perform better on trust-linked outcomes than visually identical PDPs without credentials, with e-WOM and Customer Satisfaction (CSAT) improvements emerging as downstream correlates (Spence, 1973; National Retail Federation, 2024).

Market and Policy Context: The Global Landscape and India as a Salient Testbed

India offers a compelling context needed to examine visual trust because of high-growth categories that lean heavily on imagery to set expectations at the Product Detail Page (PDP). Current forecasts project India's e-commerce market to rise from approximately USD 125 billion in FY2024 to USD 345 billion by FY2030, propelled by smartphone penetration, low-cost data, and the Unified Payments Interface (UPI). As volumes scale, small percentage shifts in return rates or post-purchase dissatisfaction cascade into material operating costs. This pattern is underscored by United States benchmarks, where the National Retail Federation (NRF) reported USD 890 billion in returns in 2024, representing 16.9% of annual sales—a structural drag on margins for online categories that lean heavily on images.

Figure 4 – Visual trust bottleneck



This visual dependency introduces specific risks across different retail segments, as expectations recalibrate around speed and visual appeal.

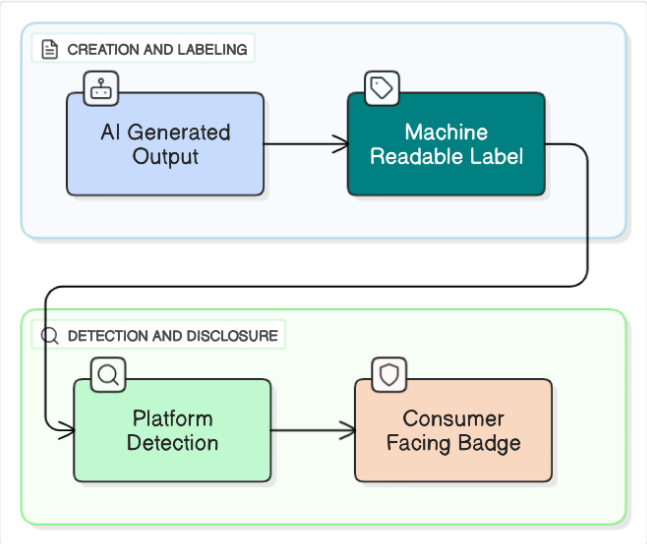
Table 2: Risk Factors in High-Velocity E-Commerce Categories in India

Category	Visual Dependency	Common Issues	Impact
Online Food Delivery	High	Over-promised imagery leading to poor presentation or packaging	High return rates, low ratings, brand switching
Quick Commerce	High	Misleading packaging visuals and unrealistic freshness cues	Customer complaints, reduced repeat usage
Apparel & Beauty	Very High	Color/fit mismatch, unrealistic model representation	High return rates, CSAT drops, negative reviews
Electronics	Moderate	Spec exaggeration, misleading product thumbnails	Increased support tickets, reduced trust in listings

Simultaneously, the regulatory trajectory across districts is converging on transparency. In the European Union, the AI Act (Regulation (EU) 2024/1689) mandates the

machine-readable marking of synthetic media, enabling platforms to detect and disclose synthetic origins to users.

Figure 5: AI content labelling workflow



In the United States, the Federal Trade Commission (FTC) continues to warn that deceptive AI tools trigger Section 5 enforcement, emphasizing there is no exemption from consumer protection laws for AI. Similarly, the UK's Advertising Standards Authority (ASA) has expanded proactive monitoring of AI-generated content, while in India, the Advertising Standards Council of India (ASCI) emphasizes truthful representation and substantiation that naturally align with provenance-backed disclosures. Architectures built on C2PA-based Content Credentials directly address these converging expectations, lowering compliance risk while improving consumer understanding.

The Verifiable Authenticity Framework

This framework restores the separating equilibria in visual communication by combining C2PA-first provenance with consumer-visible verification.

Pillar 1: Initial Adoption of Content Credentials via C2PA Technology

This pillar focuses on embedding trust directly into the images and videos used in e-commerce by leveraging the C2PA standard, which allows brands to attach secure, verifiable information to each piece of media. The operationalization of this pillar follows a sequential process.

Capture & Attest

The provenance lifecycle begins at the point of creation. When a photo or video is captured-whether in a professional studio, a vendor's kitchen, or a retail environment, C2PA-supported tools automatically embed critical metadata into the file. This attestation

includes device information (such as the camera model), the creator's cryptographic identity, a secure timestamp, and relevant location data, thus ensuring that the asset possesses a trustworthy and verifiable origin from inception.

Preserve Edit History

As the asset progresses through post-production processes, such as cropping, colour correction, or compositing, every modification must be cryptographically recorded. This requires agencies and freelancers to utilize compatible software that maintains the full edit history. Workflows that inadvertently or intentionally strip away credentials, such as saving a file without its metadata payload, are strictly prohibited to ensure the final image accurately reflects its original context and subsequent alterations.

Validate at Ingest

Upon upload to a brand's Content Management System (CMS) or Product Information Manager (PIM), the asset undergoes automated validation. Systems are configured to accept images with valid, intact credentials while flagging or quarantining assets that lack metadata or fail cryptographic checks. Logging these validation outcomes is essential for maintaining robust audit trails and ensuring that only verifiably trustworthy images progress to consumer-facing platforms.

Surface in User Experience

At the point of display on the Product Detail Page (PDP), the framework surfaces a Content Credentials badge. When a consumer interacts with this badge, a user-friendly modal opens to explain the image's origin and edit history in plain, non-technical language. This transparency empowers consumers to evaluate the authenticity of the visual representation without requiring specialized cryptographic knowledge.

Trust and Revocation

To maintain the security and integrity of the system, brands must manage a secure list of trusted signers, such as approved photographers or authorized agencies. If a signer’s credentials are ever compromised, the infrastructure supports immediate revocation, which automatically invalidates the credentials and removes the verification badge from all affected images across the platform. This mechanism ensures that compromised assets are swiftly neutralized.

Note: C2PA is open source, works with many file formats, and is moving toward international standardization (ISO/DIS 22144). It is supported by major content platforms and tools, making it a scalable and future-proof solution (Coalition for Content Provenance and Authenticity, 2025a).

Pillar 2: Consumer-Facing Transparency: The Role of Browser and App Extensions

To ensure the Verifiable Authenticity Framework is not only robust but also accessible to the average consumer, its second pillar, Verifiable Transparency, can be operationalized through user-friendly browser and app extensions. This approach shifts the burden of trust verification from the consumer to a streamlined, automated process.

A proposed mechanism is a browser extension or a similar app extension for mobile e-commerce platforms. This extension would automatically analyse the C2PA embedded within images as the consumer browses a product page. When an authentic, verifiable image is detected, the extension can render a clear, intuitive visual signal such as a green checkmark or a "Content Verified" badge-in the corner of the image.

Clicking on this badge would allow the consumer to view a simplified authenticity report, detailing key provenance information: who created the content, the date of its last modification, and the source device. This not only empowers the consumer with immediate, verifiable information at the crucial point of purchase but also serves as a strong signal that the brand is committed to transparency. By integrating this functionality directly into the user's browsing experience, firms can restore trust with minimal friction, making authenticity a seamless part of the online shopping journey. This mechanism directly translates a costly and credible C2PA signal into a simple, separating equilibrium for the consumer.

Pillar 3: Commitment to Verifiable Transparency (On-Chain Optional)

This pillar emphasizes making provenance visible and meaningful to consumers, turning back-end verification into a front-end trust signal.

For Most Products

Displaying the C2PA badge on the PDP is sufficient. It provides consumers with a clear, verifiable signal that the image is authentic and responsibly edited.
Note: This approach balances transparency with simplicity, making it suitable for everyday products.

For High-Value or Collectible Items

In categories like limited-edition artisanal sarees or collectibles, brands may choose to anchor the image’s credentials to a public blockchain. This creates a permanent, tamper-proof record that enhances trust in resale or secondary markets.
Note: On-chain notarization is optional and should be used only when justified by product value, user experience, and privacy regulations.

The Separating Effect

Credentialed images require real effort-authentic photography, secure editing, and proper validation. This creates a separating equilibrium:

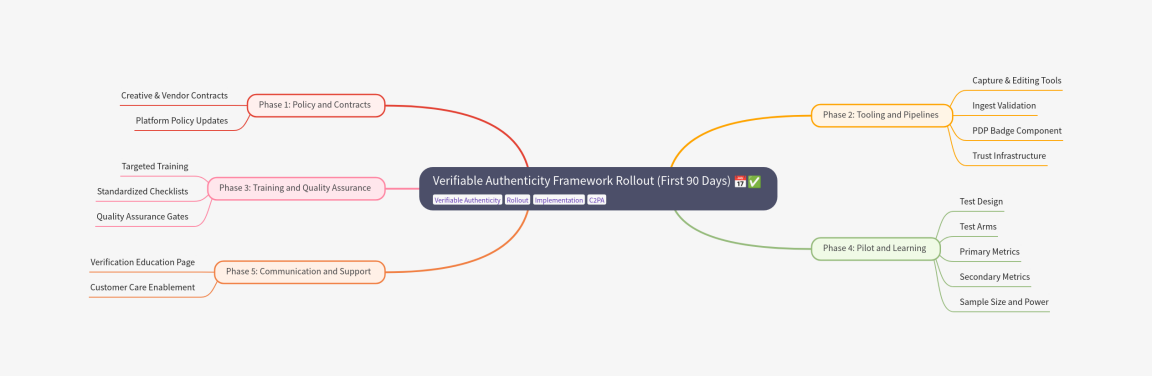
- Honest sellers can prove their integrity.
- Dishonest sellers face higher barriers and cannot easily mimic credentialed content.

Note: This restores credible differentiation in digital marketplaces, helping consumers make informed choices and reducing post-purchase dissatisfaction.

Implementation Playbook (First 90 Days)

The rollout of the Verifiable Authenticity Framework is structured into five sequential phases designed to build operational readiness, enforce provenance discipline, and establish consumer-facing credibility. Each phase addresses a distinct layer of organizational transformation-from policy alignment to consumer education.

Figure 6: Mind-map of implementation phases



Phase 1: Policy and Contracts

This foundational phase ensures that all stakeholders—internal teams, external vendors, and platform operators—are contractually aligned with the framework's requirements. Organizations must revise creative and vendor contracts to mandate the use of C2PA Content Credentials across all visual assets. These agreements should specify continuous manifest retention across the asset lifecycle, outline permissible edit types (such as cropping or colour correction), and ensure the delivery of provenance metadata alongside final assets. Simultaneously, platform policies must be updated so that PDP images lacking valid credentials are strategically excluded from promotional placements, homepage visibility, and featured listings, thereby enforcing a baseline of authenticity for high-visibility content (Coalition for Content Provenance and Authenticity, 2025a).

Phase 2: Tooling and Pipelines

Phase 2 establishes the technical infrastructure required to support credentialed content across the media supply chain. Capture and editing tools must be configured to embed and preserve C2PA manifests, supporting device-level attestation, non-destructive editing workflows, and manifest chaining across revisions. At the ingest stage, Content Management Systems (CMS) and Product Information Managers (PIM) must be upgraded to verify credentials automatically, quarantine assets that fail validation, and log outcomes for audit readiness. Furthermore, a PDP badge component must be developed using C2PA's design guidelines needed to surface credentials via a user-friendly modal. Finally, robust trust infrastructure is required to maintain a list of approved signers, execute revocation checks to detect compromised certificates, and ensure version control for manifest features.

Phase 3: Training and Quality Assurance

Operational execution requires that teams be equipped with the appropriate knowledge to sustain the framework reliably. Targeted training sessions for studios, agencies, and merchandisers should cover credential embedding, manifest retention protocols, and validation workflows. To formalize this discipline, shoot and post-production workflows must be governed by standardized checklists. Quality assurance mechanisms, including pre-publish validation gates that block non-credentialed assets and spot audits on 5% of weekly uploads, are critical to maintaining systemic.

Phase 4: Pilot and Learning

This phase assesses the real-world impact of displaying Content Credentials badges on product pages to understand their influence on consumer trust, satisfaction, and purchase decisions before a broader rollout. Utilizing a controlled A/B test, product pages randomly assigned to test groups, matched by SKU to ensure fair comparison, and clustered by geography or device to isolate the effect and avoid spillover. The test arms include a control group viewing standard images, Treatment A viewing images with a C2PA badge and modal, and Treatment B viewing the badge alongside explanatory policy text. Primary metrics focus on return rates, rating distributions, and image-related complaints, while secondary metrics track conversion rates, time on PDP, modal click-through rates, and repurchase behavior. To detect meaningful changes, such as a 1.5–2.0 percentage point drop in return rates, the test leverages historical data to estimate variance and requires tens of thousands of sessions per group over a 2 to 4 week period, adjusted for seasonal effects.

Phase 5: Communication and Support

The final phase focuses on building consumer awareness and reinforcing the brand's commitment to transparency by educating users and empowering support teams. A dedicated verification education page should be created to explain the meaning and

purpose of the Content Credentials badge, detailing what information it reveals and why it matters for product authenticity. This page should include Frequently Asked Questions (FAQs) and visual walkthroughs to normalize provenance checking as a standard part of the shopping experience. Concurrently, customer service agents must be trained to confidently explain the verification process. Updating scripts and response templates to reference the verification system enables support teams to strengthen brand credibility and ensure consistent messaging across consumer touchpoints.

Table 3: Operational Summary of Phases 2–5

Phase	Key Activities	Responsible Teams	Success Metrics
1. Policy & Contracts	Update vendor contracts; revise platform image policies.	Legal, Procurement, Brand	Signed contracts mandating C2PA; updated platform policy documents.
2. Tooling & Pipelines	Configure C2PA-enabled tools; develop PDP badge component; implement CMS validation.	IT, Engineering, Product	Functional credential validation in CMS; UI badge component ready for deployment.
3. Training & QA	Train studios, agencies, merchandisers; implement spot audits and pre-publish gates.	Creative Teams, Agencies, QA	>90% of new assets are credentialed; zero non-credentialed assets published.
4. Pilot & Learning	Launch controlled A/B test on high-traffic categories.	Data Science, Product	Empirical data on return rates, ratings, conversion, and repurchase intent.
5. Communication & Support	Launch consumer FAQ page; train customer care teams.	Marketing, Customer Support	Reduced ticket resolution time for image complaints; positive consumer sentiment.

Note: Empowering support teams with this knowledge strengthens the brand's credibility and ensures consistent messaging across consumer touchpoints.

Methods Roadmap: Evaluating the Signal

This section builds on the pilot experiment described in Phase 4 of the implementation plan and provides a formal research design to evaluate the impact of Content Credentials on consumer behaviour in digital commerce.

Research Questions and Hypotheses

This study is guided by three primary research questions regarding the impact of verification badges: whether their presence reduces product return rates, whether they

improve customer sentiment (as reflected in ratings and complaints), and whether they positively influence repurchase behaviour and electronic word-of-mouth (e-WOM). Based on the premise that verified images set accurate consumer expectations and foster post-purchase satisfaction, we propose several hypotheses. First, we hypothesize that Product Detail Pages (PDPs) displaying Content Credentials will demonstrate a 1.5–2.0 percentage point reduction in return rates compared to standard, unverified PDPs. Second, we expect these verified pages to receive fewer low-star reviews and image-related complaints. Finally, we hypothesize that verified visuals will yield modest increases in repurchase rates within a 60- to 90-day window.

Experimental Design and Randomization

The experiment employs a randomized controlled trial structure to evaluate these hypotheses. PDPs will be assigned to one of three experimental groups. To ensure a fair comparison, a PDP-level randomization strategy will be utilized by meticulously matching products via stock-keeping units (SKU). Furthermore, to minimize the risk of spillover effects between the experimental groups, the test will cluster assignments by geography or device type.

Table 4: Experiment Design

Group	Description
Control	Standard product images without any badge
Treatment A	Product images with Content Credentials badge and verification modal
Treatment B	Same as Treatment A, plus explanatory policy text

Metrics Overview

To capture the comprehensive impact of the verification badges, the evaluation will track both primary and secondary metrics. Primary metrics focus on immediate behavioural and sentiment shifts, such as return rates, rating distributions, and the volume of image-related complaints. Secondary metrics track broader engagement and loyalty indicators, including conversion rates, time spent on the PDP, modal click-through rates, add-to-cart rates, repurchase rates, and the overall polarity of e-WOM. These metrics align with the initial pilot phase but provide the granular methodological detail necessary for rigorous evaluation.

Table 5: Metrics Overview

Metric Type	Metrics
Primary	Return rate, Rating distribution, Image-related complaints
Secondary	Conversion rate, Time on PDP, Modal CTR, Add-to-cart rate, Repurchase rate, e-WOM polarity

Note: These metrics are consistent with those introduced in Phase 4 but are expanded here with methodological detail.

Sample Size and Power

To detect a meaningful change—such as a 1.5–2.0 percentage point drop in the return rate from a 10% baseline—the study requires robust statistical power. Historical data will be utilized to estimate the expected variance, dictating that the test requires tens of thousands of user sessions per experimental group. Furthermore, the experiment will run for a duration of two to four weeks, with the timing carefully adjusted to avoid seasonal anomalies, such as holiday shopping spikes, thereby ensuring the results are reliable and are not driven by random chance or external market events.

Analysis Plan

To isolate the causal impact of Content Credentials on consumer behavior, the evaluation will employ a difference-in-means approach to compare baseline return rates and Customer Satisfaction (CSAT) scores across the Control, Treatment A, and Treatment B groups. Additionally, a logistic regression model will be utilized to analyze binary outcomes, such as the likelihood of a product return or the incidence of an image-related complaint, allowing researchers to control external variables including promotional events and seasonality. For practitioner relevance, a "meaningful effect" is defined a priori as a statistically significant 1.5–2.0 percentage point reduction in return rates. This threshold represents a material shift in consumer behavior that would directly translate into operational cost savings, thereby justifying the infrastructure investment required for C2PA integration.

Measurement and Integrity

To ensure accurate tracking, system reliability, and a seamless user experience, the Content Credentials Badge system incorporates several key mechanisms. These safeguards work together to maintain trust, optimize performance, and ensure the integrity of content verification.

Table 6: Measurement and Integrity

Feature	Purpose	Behavior	Impact on User Experience
Interaction Logging	Track user engagement with the badge and modal	Every click on the badge or modal open is recorded	Enables analytics and usage insights without affecting performance
Validation Tracking	Monitor image credential verification outcomes	Records whether the image passes or fails credential checks	Supports audit trails and system reliability
Caching	Improve performance and reduce latency	Temporarily stores verification results to avoid repeated checks	Ensures fast and smooth page load
Silent Badge Removal	Prevent user confusion from failed validations	Badge is removed quietly if validation fails (e.g., hash mismatch or revoked certificate)	Maintains trust and avoids alarming users

Each of these mechanisms plays a distinct role:

- **Interaction Logging** enables detailed engagement analytics and supports system diagnostics without impacting performance.
- **Validation Tracking** ensures traceability by recording the outcome of each credential check, whether successful or failed.
- **Caching** improves efficiency by storing verification results temporarily, reducing redundant checks and maintaining fast page loads.
- **Silent Badge Removal** protects user trust by discreetly removing badges that fail validation, avoiding confusion or alarm.

Together, these features ensure that technical failures do not disrupt the user experience while preserving the transparency and reliability of the content verification system.

Threats to Validity

To guarantee the integrity and reliability of the experiment's outcomes, multiple safeguards have been implemented needed to minimize the impact of external factors and reduce the risk of manipulation.

Table 7: Threats to Validity

Threat	Mitigation Strategy	Purpose & Impact
Seasonality & Promotions	Stagger testing over time and flag promotional events during analysis	Prevents skewed results due to time-based or marketing-driven fluctuations
Spillover Effects	Randomize at the cluster level (e.g., region or device)	Ensures users don't encounter multiple test versions, preserving test isolation
Adversarial Behavior	- Validate credentials server-side - Block untrusted image sources - Investigate sudden validation changes	Protects against badge misuse or forgery by low-quality sellers

These measures collectively protect the integrity of the experiment, ensuring that findings reflect genuine user behavior and system performance without interference from seasonal trends, cross-condition contamination, or malicious actors.

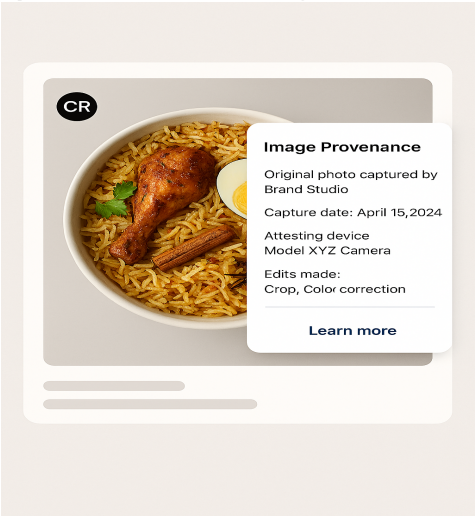
Consumer Psychology and UX: Making Credentials Legible

For Content Credentials to be effective, they must be both technically robust and intuitively understandable. This section explores how principles from consumer psychology and user experience (UX) design shape the success of verification badges in digital commerce.

Simplicity Over Complexity

Consumers make rapid decisions when shopping online and rarely engage with technical explanations, especially those involving cryptographic terms or complex metadata structures. Therefore, the badge and modal system must intuitively answer fundamental questions, such as where an image originated and whether it has been edited. Rather than serving merely as a decorative element, the badge must function as a practical tool for shoppers to verify visual authenticity. This streamlined approach builds trust without overwhelming the user with unnecessary cognitive load.

Figure 7: Mock-up of Consumer-Facing Content Credentials Badge



Clear and Familiar Design

To encourage interaction and build consumer confidence, the verification interface should rely on familiar design patterns. The badge itself must be clearly clickable, prompting a modal that behaves consistently across all product pages and platforms. Within this modal, plain language is essential; it should explain the image's origin and any subsequent edits in simple, jargon-free terms. For instance, a clear explanation such as, "Original photo captured by Brand Studio on March 12, 2025; edits: crop, colour correction," effectively builds trust and encourages users to routinely engage with provenance data.

Localization and Inclusivity

Given that India's digital marketplace is highly diverse and multilingual, accessibility is a critical component of UX design. Key phrases within the verification modal should be translated concisely into regional languages to ensure broad comprehension. Furthermore, the tone of the information should remain neutral and informative, avoiding promotional language or technical jargon. This localized approach ensures that all users, regardless of their primary language or digital literacy level, can fully understand and trust the verification system.

Normalizing Provenance Checks

Just as modern consumers routinely consult seller ratings or customer reviews, provenance checking has the potential to become a standard, habitual part of the digital shopping experience. Over time, as users begin to expect verification badges on product images, platforms can reinforce this behaviour by strategically highlighting verified visuals in search results or promotional banners. Research in consumer psychology supports this approach, demonstrating that clear visual cues can recalibrate consumer expectations and significantly reduce impulse-driven post-purchase disappointment (Rodrigues et al., 2021).

Table 8: UX principles and implementation strategies

UX Principle	Design Strategy	User Benefit
Simplicity Over Complexity	Use plain language; avoid cryptographic jargon	Reduces cognitive load; encourages interaction
Clear and Familiar Design	Clickable badge; consistent placement; modal with intuitive info	Builds trust; improves usability across platforms
Localization and Inclusivity	Translate key phrases; use neutral tone	Ensures accessibility for diverse, multilingual users
Normalizing Provenance Checks	Highlight verified images in search and banners	Encourages habitual verification behavior

This integrated approach ensures that Content Credentials are not only secure and verifiable but also legible, inclusive, and behaviorally aligned with how users shop and make decisions online.

Cost–Benefit Logic for Managers

Implementing Content Credentials and provenance verification requires thoughtful investment, but long-term returns can significantly enhance business performance, customer trust, and regulatory resilience.

Cost Considerations

Organizations should anticipate several upfront and operational costs. Investment in tooling, such as software and verification systems that support C2PA standards, is necessary to enable credential embedding. Additionally, teams including photographers, editors, and merchandisers must undergo training to validate credentials correctly, ensuring consistency and reducing errors. Workflow updates will also be required to adjust existing asset management systems and to publish pipelines needed to accommodate credential preservation. Finally, organizations may experience minor operational delays during the transition, such as rescheduled shoots or product launches if credentials fail validation; however, these delays are manageable and can be minimized with a phased rollout approach.

Expected Benefits

Conversely, the long-term advantages of verifiable provenance extend across multiple business functions. Verified images help set accurate consumer expectations, directly contributing to lower return rates and improved inventory turnover. This transparency also boosts Customer Satisfaction (CSAT) and brand loyalty, as consumers feel more confident in the authenticity of their purchases. Operationally, fewer complaints about misleading visuals result in a reduced volume of support tickets, freeing up customer service resources. Furthermore, credentialed content provides regulatory flexibility by aligning with emerging transparency laws, thereby reducing legal exposure, and improving

audit readiness. Ultimately, brands that adopt transparent practices achieve significant reputational gains, positioning themselves as trustworthy and forward-thinking in a competitive marketplace.

Table 9: Cost–Benefit Summary

Category	Details	Impact
Tooling	Software upgrades for C2PA compliance	Enables credential embedding and verification
Training	Upskilling teams on credentialing workflows	Reduces errors and improves consistency
Workflow Updates	Adjusting asset pipelines to support credentials	Ensures credentials are preserved across systems
Operational Delays	Occasional rescheduling due to credential issues	Minor disruptions; manageable with phased rollout
Lower Return Rates	Accurate visuals reduce buyer remorse	Saves cost and improves inventory turnover
Improved CSAT	Authentic images boost consumer confidence	Enhances brand loyalty and satisfaction
Fewer Support Tickets	Less confusion over visuals reduces complaints	Frees up customer service resources
Regulatory Flexibility	Compliance with transparency laws	Reduces legal risk and improves audit readiness
Reputational Gains	Transparent practices build brand trust	Strengthens market position and consumer perception

Rollout Strategy

To maximize impact and minimize disruption, brands should begin implementation in high-return categories where visual dependency is critical. For instance, apparel requires visual accuracy for sizing, fit, and color expectations, while beauty products rely heavily on images to convey texture and finish. Similarly, prepared foods depend on authentic visuals to communicate freshness and portion size. Because these categories are highly vulnerable to consumer disappointment when images overpromise, they represent ideal candidates for early adoption.

Boundary Conditions and Operational Risks

While the Verifiable Authenticity Framework provides a robust mechanism for restoring visual trust, its implementation is subject to several boundary conditions and practical barriers. First, the framework's efficacy is highly dependent on broader ecosystem adoption; if major hardware manufacturers, editing software suites, and downstream platforms lag in native C2PA integration, the operational burden disproportionately falls on individual brands. Second, there are inherent risks of credential stripping that is both malicious and accidental, and specifically where social sharing platforms or unoptimized

content delivery networks (CDNs) compress images and inadvertently remove critical provenance metadata. Third, seller heterogeneity presents a structural challenge. Smaller or less technologically sophisticated merchants may experience higher user experience (UX) friction and initial integration costs, potentially creating an unbalanced marketplace where only well-resourced sellers can afford to signal authenticity. Finally, as organizations capture device-level and location-based attestation, they must navigate strict data privacy considerations, ensuring that provenance metadata does not inadvertently expose sensitive creator or vendor information to the public.

Limitations and Future Research

While the Verifiable Authenticity Framework presents a promising approach to restoring visual trust in digital commerce, it currently remains conceptual. Because the framework is built on theoretical foundations and initial implementation assumptions rather than executed empirical data, its practical efficacy must be rigorously validated. Future research should prioritize translating this design into applied settings across several key dimensions.

Empirical and Contextual Validation

First, researchers should execute robust A/B tests across diverse product categories and seasonal shopping events. Consumer expectations for apparel fundamentally differ from those for electronics, and purchasing behaviour often shifts during high-volume periods like festive sales. Testing the framework across these contexts, as well as piloting it in varied geographic markets outside of India (such as Southeast Asia, Europe, and Latin America), will determine whether the benefits of Content Credentials are universally applicable or are highly context dependent.

Consumer Psychology and UX Innovation

Second, future studies must examine consumer comprehension and behavioural responses to provenance signals. Even a technically sound verification system will fail if users do not understand or trust the interface. Research should investigate how different demographic groups varying by age, region, and digital literacy interpret the badge and modal to guide inclusive design improvements. Furthermore, behavioural experiments should explore how these trust signals influence impulse buying and long-term brand loyalty, while UX innovation studies could test alternative formats, such as animations or audio cues, to maximize user engagement.

Economic Modelling and Security Monitoring

Finally, decision-makers require clear financial and security models to justify infrastructural investments. Future research should develop economic models that contrast implementation costs, such as new tooling and staff training, with operational

savings derived from reduced return rates and customer support tickets. Concurrently, technical studies must monitor adversarial behaviour in production environments, tracking the frequency of credential stripping or spoofing attempts. Evaluating the real-world effectiveness of server-side validation and revocation checks will provide critical feedback for refining the C2PA standard and aligning future iterations with evolving global transparency policies.

Additional Opportunities to Enhance the Research

Beyond foundational validation and economic modeling, there are several promising avenues to deepen the academic and practical understanding of visual trust mechanisms. First, controlled behavioral experiments should be conducted to explore how cryptographic trust signals influence complex consumer dynamics, such as impulse buying, overall brand perception, and long-term customer loyalty. Second, UX innovation studies could evaluate alternative badge formats; these may include subtle animations or accessible voice cues needed to optimize consumer engagement without introducing purchase friction. From a technical perspective, AI integration presents an opportunity to deploy machine learning algorithms that predict which specific products, retail categories, or consumer segments benefit the most from credentialing interventions. Finally, continuous policy collaboration is essential; researchers should actively partner with regulatory bodies to ensure that emerging empirical findings both inform, and remain aligned with, the rapid evolution of global digital transparency laws.

Conclusion: Rebuilding Trust in the Age of Synthetic Media

In today’s digital marketplace, especially in fast-growing regions like India product visuals play a pivotal role in shaping consumer expectations. Whether it is a luxury saree, a skincare serum, or a plate of Hyderabadi Biryani, shoppers rely heavily on images to inform their purchase decisions. However, with the rise of generative AI and low-cost image editing tools, it has become easier than ever to create visually polished yet misleading content. This dynamic has led to a crisis of visual trust, where consumers can no longer be confident that what they see is what they will ultimately receive.

To address this challenge, this article proposes the Verifiable Authenticity Framework, built on the dual pillars of C2PA Content Credentials and consumer-visible verification. By embedding secure, tamper-evident metadata into visual assets, brands can cryptographically prove who created the content, when it was captured, and what edits were subsequently made. This server-side validated metadata is then surfaced to consumers via a clear, plain-language verification badge and modal on the Product Detail Page. Instead of blindly trusting marketing claims, shoppers can intuitively answer questions regarding the image's authenticity, thereby empowering them to "see for themselves."

Circling back to the initial example of the Hyderabadi Biryani, this framework directly mitigates the expectation-reality mismatch. Rather than encountering a hyper-realistic

image that results in a disappointing delivery, a consumer would interact with a Content Credentials badge on the listing. The modal would transparently reveal that the original photo was captured by the brand's verified studio on a specific date, noting only permissible edits like cropping or colour correction. If the image had been heavily retouched or sourced from a stock library, the credentials would reflect that reality, or the badge would be silently removed if validation failed. This level of transparency empowers the shopper to make informed decisions and restores trust in visual representations.

For international managers and DBA practitioners, this proposed framework and evaluation design offer a testable, scalable strategy to address a critical operational challenge. By moving from mere promises to cryptographically backed proof, brands can restore visual credibility, reduce costly return rates driven by misaligned expectations, and build long-term consumer loyalty. Furthermore, proactive adoption aligns organizations with emerging cross-jurisdictional regulations on digital content provenance, differentiating them in a crowded, AI-driven marketplace. India’s digital commerce environment that is characterized by mobile-first users, multilingual diversity, and visual-heavy shopping behaviour, which in turn can serve as an ideal testbed for this shift, ultimately setting a new global standard for market integrity in the age of synthetic media.

Acknowledgements

The author thanks colleagues and reviewers for their constructive feedback during the development of this manuscript.
Author’s Note on AI Assistance: During manuscript preparation, the author used generative AI tools only for language polishing and figure drafting. Prompts and edits are available upon request. All outputs were reviewed and validated by the author, and no confidential or proprietary data were provided to AI tools.
No external funding was received for this work. The author declares no conflicts of interest.

References

Adobe Inc. (2025, June 18). *Content Credentials overview*. Adobe Help Center. https://helpx.adobe.com/creative-cloud/help/content-credentials.html

Advertising Standards Authority. (2025, February 7). *AI, advertising, and the policy landscape – CAP proactive monitoring*. https://www.asa.org.uk/news/ai-advertising-and-the-policy-landscape-cap-proactive-monitoring.html

Advertising Standards Council of India. (2024). *Guidelines for advertisements making environmental/green claims*. https://www.ascionline.in/wp-content/uploads/2024/01/Guidelines-for-Advertisements-Making-Environmental-Green-Claims.pdf

Advertising Standards Council of India. (n.d.). *The ASCI code*. https://www.ascionline.in/the-asci-code/

Burges Salmon. (2025, May 6). *Advertising and AI – ASA and CAP publish 2024 report highlighting how they use AI*. https://www.burges-salmon.com/articles/102ka6v/advertising-and-ai-asa-and-cap-publish-2024-report-highlighting-how-they-use-ai/

Coalition for Content Provenance and Authenticity. (2025a). *Content Credentials: C2PA technical specification (Version 2.1)*. https://spec.c2pa.org/specifications/specifications/2.1/specs/C2PA_Specification.html

Coalition for Content Provenance and Authenticity. (2025b). *C2PA implementation guide for content ecosystems*. https://www.c2pa.org/resources

Content Credentials. (2025). *Content Credentials* [Website]. https://contentcredentials.org

Davis Polk. (2024, October 3). *FTC announces new enforcement initiative targeting deceptive AI practices*. https://www.davispolk.com/insights/client-update/ftc-announces-new-enforcement-initiative-targeting-deceptive-ai-practices

Digimarc. (2024). *C2PA 2.1: Strengthening Content Credentials with digital watermarks*. https://www.digimarc.com/blog/c2pa-21-strengthening-content-credentials-digital-watermarks

DLA Piper. (2025, August 7). *Latest wave of obligations under the EU AI Act take effect: Key considerations*. https://www.dlapiper.com/en-hk/insights/publications/2025/08/latest-wave-of-obligations-under-the-eu-ai-act-take-effect

European Parliament. (2025). *Answer to parliamentary question on AI Act transparency obligations (Article 50)*. https://www.europarl.europa.eu/doceo/document/E-10-2025-001920-ASW_EN.pdf

Ghosh, K., Bellinger, C., Corizzo, R., Branco, P., Krawczyk, B., & Japkowicz, N. (2024). The class imbalance problem in deep learning. *Machine Learning*, *113*, 4845–4901. https://doi.org/10.1007/s10994-022-06268-8

India Brand Equity Foundation. (2025). *E-commerce industry in India*. https://www.ibef.org/industry/ecommerce

International Organization for Standardization. (2025). ISO/DIS 22144: Authenticity of information Content credentials. https://www.iso.org

National Retail Federation. (2024, December 5). *NRF and Happy Returns report: 2024 retail returns total $890 billion*. https://nrf.com/media-center/press-releases/nrf-and-happy-returns-report-2024-retail-returns-total-890-billion

Rodrigues, R. I., Lopes, P., & Varela, M. (2021). Factors affecting impulse buying behavior of consumers. *Frontiers in Psychology*, *12*, 697080. https://doi.org/10.3389/fpsyg.2021.697080

Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics*, *87*(3), 355–374. https://doi.org/10.2307/1882010

Statista. (2025). *Food delivery industry in India – Statistics & facts*. https://www.statista.com/topics/7741/food-delivery-industry-in-india/

Tolosana, R., Vera-Rodríguez, R., Fierrez, J., Morales, A., & Ortega-García, J. (2020). Deepfakes and beyond: A survey of face manipulation and fake detection. *Information Fusion*, *64*, 131–148. https://doi.org/10.1016/j.inffus.2020.06.014