EMERGING LEGAL LANDSCAPE FOR GENERATIVE AI ACROSS DIFFERENT REGIONS

SILVIA A. CARRETTA, MARILENA HYERACI,
VIJAY MAUREE AND ALESSANDRA SALA

POLICY PAPER
Edition 1

TABLE OF CONTENTS

	Executive Summary	03
	AI and Multimedia Authenticity Standards Collaboration (AMAS)	03
	Key takeaways	03
	Introduction	05
1	**Transparency obligations**	06
2	**Mandatory Labeling: Implementing global disclosure standards for AI-generated content**	09
3	**Enforcement provisions: Analyzing rapid-response protocols for taking down harmful synthetic media**	12
4	**Media literacy**	15
	Platform providers and AI intermediaries	
	AI developers and deployers	
5	**Copyright & Training: Navigating “fair use” in the era of massive AI data scraping**	17
6	**Creator Compensation: How can copyright holders exercise their rights?**	19

EXECUTIVE SUMMARY

This paper analyzes emerging legal and policy approaches to deepfakes, misinformation/disinformation, and multimedia authenticity from Africa, Asia, Middle East, Europe and Americas regions. The policy interventions from the following countries in these regions are considered in the paper: Brazil, China, the European Union (EU), India, Singapore, and the United States. While definitions and legal provisions vary by jurisdiction (including intent, harm thresholds, and context), the core regulatory direction is broadly converging around transparency/provenance obligations, copyright and training, creator compensation, risk-based restrictions on high-harm uses and enforcement provisions.

AI and Multimedia Authenticity Standards Collaboration (AMAS)

The AI and Multimedia Authenticity Standards Collaboration is a global initiative advancing standardization in the rapidly evolving field of AI-generated and altered media. By identifying gaps and driving the development of new standards, we support transparent, privacy-conscious, and rights-respecting practices. Our work also aims at informing policy and regulatory frameworks to promote legal compliance and safeguard public trust.

Led by the World Standards Cooperation, the collaboration serves as a vital forum for dialogue among standards developers, civil society organizations, technology companies, and other key players.

Convened by ITU under the auspices of the World Standards Cooperation, the collaboration was launched at the AI for Good Global Summit in 2024.

Learn more at **https://aiforgood.itu.int/multimedia-authenticity/**, including a full list of members, or contact the Secretariat at **amas-secretariat@itu.int**

KEY TAKEAWAYS

- Transparency is becoming the baseline rule: Users increasingly have a right to know when content is AI-generated or manipulated. The EU's AI Act (Article 50) anchors this trend through machine-readable marking obligations for providers and disclosure obligations for deployers.
- The global regulatory landscape on deepfakes can currently be framed around three macro-categories: (1) Prevention (takedowns, evidentiary rules, platform liability), (2) Disclosure (labeling and transparency obligations), and (3) Accountability (enforcement for privacy, copyright, or criminal violations).
- Jurisdictions diverge on how mandatory and how prescriptive labeling must be: China's deep-synthesis framework is comparatively stringent and treats transparency as a mandatory requirement, with limited expressive carve-outs.
- Elections are a primary high-risk context: Brazil's clearest binding disclosure rule is electoral (TSE Resolution No. 23.732/2024), and Singapore imposes a tightly defined election-period deepfake ban while requiring disclosures outside that window for election campaigning.

- Platform obligations are expanding from reactive takedowns to proactive risk management: The EU's Digital Services Act (DSA) imposes systemic risk assessment and mitigation duties on very large platforms (VLOPs), and enforcement activity has escalated in early 2026.

- Privacy and consent have become a central enforcement lever: Regulators increasingly reframe face/voice likeness and biometric attributes as protected personal data, targeting non-consensual generation or use without needing to litigate "truth."

- Speech and satire remain the hardest boundary: Some regimes explicitly preserve space for evidently artistic/satirical works (e.g., EU AI Act exceptions), while U.S. courts have actively scrutinized and sometimes struck down broad election-deepfake bans and rigid disclaimer mandates as unconstitutional restrictions on protected expression.

- Non-consensual intimate imagery (NCII) is treated as a "no gray area" category: Across approaches discussed, sexually explicit synthetic content without consent is consistently treated as a severe harm category demanding rapid removal and strong remedies.

- A summary of the legislation framework in the different countries is shown in the table below:

Country/Region	Standalone Deepfake/ AI Law?	Primary Mechanism	Labeling / Watermarking
EU	Yes (AI Act)	Strict provider and deployer obligations	Mandatory (Machine-readable & visible)
Brazil	No (Pending PL 2338)	Electoral court bans on campaign deepfakes	Mandatory for election materials
USA	No (State-by-state laws)	Federal notice-and-takedown for intimate content	Disclaimers required in ~28 states for politics
China	Yes (CAC Measures)	Administrative control, platform suspension, criminal trials	Mandatory (Machine-readable & visible)
UAE	No (Cybercrime framework)	Penalizes reputational damage and biometric theft	Voluntary / Platform dependent
Singapore	No (Targeted Acts)	Direct takedown orders via Online Safety Commission	Voluntary (Encouraged via IMDA codes)
India	Yes (IT Rules Intermediary Regime)	2 to 3-hour takedown window or lose safe harbor	Mandatory (10% asset space rule)
Kenya	No (Pending AI Bill)	Proposed AI Commissioner & criminal prosecution	Mandatory disclosure clauses under draft
South Africa	No	Cybercrimes Act (simulated content provisions) & POPIA	Voluntary / Platform dependent

- Unlike U.S. law, EU law does not provide an open-ended "fair use" defense: AI-scraping of protected contents relies on closed, purpose-bound exceptions, most notably the text-and-data mining (TDM) exception under the Copyright Directive, remaining fact-sensitive and legally unsettled pending further case law and standardizations, particularly on opt-out mechanisms for rightsholders.

- Creator compensation may be framed as license-fee or damage claims in-court or out-of-court asserting challenges to TDM applicability, thereby making proactive licensing (including through collective licensing frameworks) a key risk-mitigation pathway alongside compliance with transparency requirements under the EU AI Act.

INTRODUCTION

The rapid proliferation of generative artificial intelligence has fundamentally disrupted how digital content is created, distributed, and consumed. From hyper-realistic deepfakes to AI-synthesized text, audio, and imagery, the line between authentic and machine-generated media has become increasingly difficult to discern. Deepfakes can become a significant threat when used to distribute fake media and information, damage reputations (of individuals, companies or countries), fabricate evidence, and defame, among other malicious and harmful actions. This technological shift poses profound challenges for governments, lawmakers, platforms, and civil society alike, touching on issues as diverse as intellectual property protection, democratic integrity, individual privacy, and the very fabric of public trust.

This document provides a concise regulatory map charting the current landscape of multimedia authenticity across countries in Africa, Asia, Middle East, Europe and Americas regions. The countries that are covered in this paper are Kenya, South Africa, Brazil, United States, China, Singapore, India and the European Union. It is structured around four foundational pillars that collectively define the emerging global framework for governing AI-generated content:

1. **Transparency obligations** – comparing the responsibilities of AI developers, AI deployers, platform providers and intermediaries across the different regions to ensure people know whether the content is AI generated or not. Transparency obligations may arise indirectly through election law, online safety law, misinformation law, data protection law, cybercrime law or platform/intermediary rules;

2. **Mandatory Labeling** - analyzing the diverse approaches to disclosure standards and provenance marking for AI-generated content, from the EU's interoperable metadata frameworks to China's state-audited dual-labeling regime;

3. **Enforcement provisions** - comparing the rapid-response protocols and legal mechanisms that jurisdictions have established for the takedown and remediation of harmful synthetic media;

4. **Copyright & Training** - examining how jurisdictions are grappling with the application of "fair use" doctrines and related exceptions to the mass ingestion of copyrighted works for AI model training, and

5. **Creator Compensation** - surveying the mechanisms through which copyright holders and content creators can assert their rights and seek remuneration in the AI era.

By mapping these five dimensions across the European Union, Italy, the United States, China, India, Singapore, Brazil, UAE, Kenya and South Africa, this document aims to provide a practical, high-level - though not exhaustive - reference for policymakers and governments to understand the evolving, and often divergent, regulatory responses to the main challenges of digital authenticity in the age of generative AI.

Section **01**

This paper is intended as a comparative policy overview as of the publication date and cannot be considered as a substitute for jurisdiction-specific legal advice. It focuses on the jurisdictions mentioned above and does not purport to provide a complete survey of all potentially relevant national laws, enforcement practices, private rights of action, or conflict-of-laws issues.

TRANSPARENCY OBLIGATIONS

Brazil: There is no broad AI developer transparency scheme. During elections AI deployers must explicitly inform the public when synthetic multimedia content generated by AI has been used or manipulated and must identify the technology used. Platform providers must adopt and publicize measures to prevent or reduce circulation of notoriously false or seriously out-of-context content affecting electoral integrity; they may face joint liability if they fail to act in election contexts. Application providers are treated as AI intermediaries during elections and must reduce circulation of electoral disinformation and act quickly when notified or ordered.

China: AI content developers must apply explicit and/or implicit labels to AI-generated synthetic content; implicit labels include metadata or digital-watermark-type identifiers. AI deployers must not delete, tamper with, fabricate or conceal AI labels; deployer-type actors are therefore subject to anti-tampering and dissemination transparency duties. Online platforms must detect, reinforce and propagate AI-generated-content labels; platforms must categorize content as confirmed, possible or suspected AI-generated and embed metadata indicating content nature and platform information. AI intermediaries must preserve, verify and propagate explicit/implicit AI labels and prevent label tampering.

European Union: AI Act Article 50 imposes AI provider/deployer disclosure obligations for synthetic content and deepfakes; Digital Services Act (DSA) adds AI platform and AI intermediary risk mitigation and notice-and-action duties. Deployers of AI systems generating or manipulating image, audio or video content constituting a deepfake must disclose that the content has been artificially generated or manipulated; deployers of certain AI-generated text informing the public on matters of public interest must also disclose AI generation. In addition, The EU AI Act Article 4, mandates that platform providers and deployers of AI systems ensure a “sufficient level of AI literacy” for their staff and users. This requirement is crucial because it promotes responsible AI use, strengthens compliance, and builds user trust, making AI literacy an essential component of media literacy.

Taking the Italian legislation as an example, we can see how Italy’s AI Law No. 132/2025 adds national sectoral governance and criminal-law provision to the provision of Article 50 EU AI Act that introduces transparency obligations for AI developers. Italy applies EU AI Act deployer disclosure rules and adds national AI law provisions, including transparency and human oversight safeguards. Platforms operating in Italy are subject to EU-level transparency and notice-and-action duties, while Italy adds criminal exposure for harmful AI-generated/manipulated content. AI intermediaries in Italy are covered by the EU DSA/AI Act baseline; Italy’s AI Law No. 132/2025 adds national enforcement and criminal-law safeguards, including against harmful AI-generated/manipulated content.

India: No developer level AI-generated content marking obligation is required in the legislation. Deployers are indirectly governed through platform terms and IT Rules obligations; platforms are required to inform users not to host impersonation, misinformation or deepfake content and to enforce those terms. Platforms are expected to identify misinformation/deepfakes and remove reported content within prescribed timelines. Intermediaries are the primary regulatory target: they must apply due diligence, notify users of prohibited content, remove reported deepfake/ misinformation content, and risk losing safe harbour for non-compliance.

Singapore: No developer level AI-generated content marking obligation is required in the legislation. Persons communicating false statements of fact may receive correction directions under POFMA. Election law also prohibits digitally manipulated candidate content during election periods and allows corrective directions. Intermediaries and platforms can receive correction or targeted correction directions under POFMA; election deepfake rules allow corrective actions such as takedown or disabling access for Singapore users during election periods. Intermediaries can be directed to disable access or comply with targeted correction directions. OCHA can require online service providers to disable access to deepfake images/videos linked to suspected criminal harms.

United States: No comprehensive federal developer transparency duty exists for AI-generated deepfakes. Developer obligations are fragmented and may arise indirectly through state laws, sectoral obligations, litigation or platform policies. Platforms and intermediaries under the TAKE IT DOWN Act must implement removal processes for covered AI-generated deepfakes and misinformation.

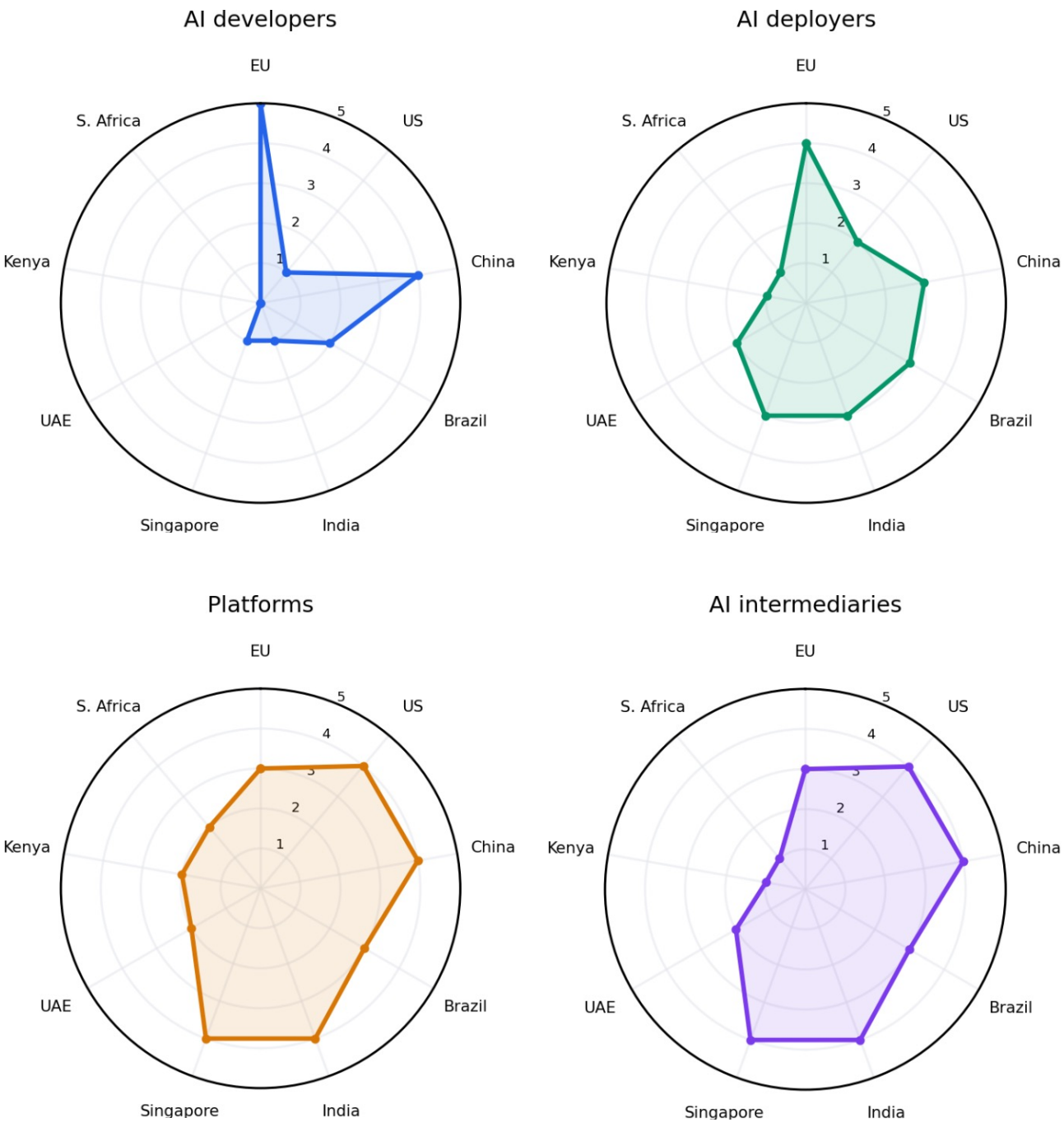
Kenya: No AI developer-specific transparency obligation for AI-generated deepfakes. Deployers/ controllers processing biometric data, including voice recognition, must comply with data protection principles and lawful-basis/consent requirements. There is no general AI-deepfake disclosure duty for deployers. Current intermediary transparency obligations for AI deepfakes are limited; data protection duties apply to controllers/processors, while proposed AI/deepfake bills may introduce more direct obligations.

South Africa: No AI developer-specific transparency obligation for AI-generated deepfakes. Deployers may face liability under existing laws if deepfakes involve unlawful disclosure of intimate images, impersonation, fraud, defamation, privacy violations or biometric data misuse, but there is no general AI-deepfake disclosure duty. Intermediary obligations are partial and arise through existing cybercrime, reporting or court-order mechanisms rather than a dedicated AI legislation.

UAE: No AI developer-specific transparency obligation for AI-generated deepfakes. Deployers may face criminal liability for disseminating rumours, fake news or false information through IT systems, but the framework is offence-based rather than a transparency/labelling regime. Platform-related obligations arise through the cybercrime/fake news framework and online content compliance expectations; no specific AI-generated content label, watermark or provenance obligation is identified. Intermediary obligations are primarily offence/compliance based under cybercrime and online content rules.

Infographic: Actor Responsibilities in Legal Frameworks for Deepfakes, Misinformation and Disinformation

Countries: US, China, Brazil, Europe (EU), India, Singapore, UAE, Kenya, South Africa | Score: 0 = absent; 5 = comprehensive



Section **02**

MANDATORY LABELING: IMPLEMENTING GLOBAL DISCLOSURE STANDARDS FOR AI-GENERATED CONTENT

Brazil: There is no general legal requirement comparable to the EU's machine-readable marking rules or China's mandatory metadata regime, as well as there is no requirement for platforms to embed provenance metadata, digital signatures, or C2PA-style traceability for all AI-generated media. The clearest binding disclosure rule applies in the electoral context, where TSE Resolution No. 23.732/2024 provides that AI-generated content in electoral propaganda must be disclosed explicitly, prominently, and accessibly, with identification of the technology used – providing a form of traceability and explainability in campaign communications. Restrictions on bots and simulated dialogue with real candidates also apply. Under Marco Civil, in addition, connection providers must keep connection logs and certain application providers must retain access logs (subject to judicial process), providing forensic traceability for investigations into coordinated disinformation or abusive deepfake dissemination. For third-party intermediaries, the key current legal traceability duties are more likely to arise from log retention, judicial cooperation, and compliance with lawful orders, rather than from authenticity metadata obligations. A bill (PL 2630/2020) creating a broader transparency regime for social networks and messaging services, especially around sponsored content, inauthentic behavior, and provider accountability, is still under discussion.

China: China has the most stringent approach globally, under the Measures for the Administration of Deep Synthesis Internet Information Services (Labelling Rules), in force since September 2025. Such rules require mandatory "dual labeling" (visible/audible labels plus embedded metadata). Prescriptive formatting requirements for explicit labels apply (e.g., minimum size/placement for image labels, prescribed approaches for audio disclosures), designed to be hard to miss and hard to remove. There is no broad exemption for satire or art. Anti-tampering rules strictly prohibit deletion, alteration, or concealment of labels. In addition, China does not rely on open standards such as C2PA; instead, it implements a state-audited metadata regime requiring embedded metadata containing the service provider's name, unique content ID, and user's digital footprint. Platforms must read uploaded metadata, add their own company identifier and distribution metadata, and lock it down. Providing tools to maliciously delete or alter tracking metadata is a criminal offense.

European Union: Under Article 50 of the AI Act (applicable from 2 August 2026), AI system providers must ensure synthetic audio, image, video or text contents are marked in a machine-readable format. Deployers must disclose to viewers when content is a deepfake. A lighter disclosure requirement applies to evidently artistic, creative, or satirical works. The Commission is working toward a code of practice (second draft published 3 March 2026) and guidelines to clarify these obligations. The draft Code of Practice advocates a "multi-layered" marking approach combining digitally signed metadata with imperceptible watermarking interwoven within the content, recognizing that no single marking technique currently meets the AI Act's requirements for effectiveness, interoperability, robustness, and reliability. Pending finalization, however, the draft Code should be treated as emerging compliance guidance rather than as a settled technical safe harbor. The Code also promotes open standards and interoperability and encourages signatories to support international and European standardization efforts for content provenance marking.

To exemplify one EU member state,Italy does not have dedicated national legislation governing the labeling of AI-generated content. As a Member State of the European Union, Italy is subject to the AI Act, and the labeling obligations set forth in Article 50 thereof apply directly. Accordingly, providers and deployers operating in Italy must comply with the EU-wide requirements for marking synthetic content and disclosing deepfakes.

India: IT Rules Amendments enacted by the Ministry of Electronics and Information Technology in February 2026, targeting "Synthetically Generated Information" (SGI), placed transparency obligations on social media intermediaries in relation to synthetic contents. The "10% rule" requires visible watermarks covering at least 10% of the visual display area. For audio contents, disclosure must be audible during the first 10% of the track. Platforms must prompt users to declare whether content is synthetically generated and deploy automated detection tools to verify declarations. Also, to be lawful, AI-generated video and audio contents must embed permanent metadata with a unique identifier tied to the platform used to create the content. Platforms cannot strip metadata and must ensure it remains intact and readable upon distribution. Unlike the EU or U.S., India does not refer to open standards, such as C2PA, in relation to the metadata regime.

Singapore: Singapore does not have a single, overarching law in relation to AI and synthetic contents. Instead, it rapidly updates existing laws to ban specific AI risks as they emerge, particularly regarding political stability and social harmony. In particular, under the Parliamentary Elections Act, AI-generated deepfakes of candidates are strictly banned during the immediate election period; outside that window, deepfakes used for campaigning must be clearly labeled as AI-generated virtual content. Under the Protection from Online Falsehoods and Manipulation Act (POFMA), ministers may issue "Correction Directions" requiring platforms to attach highly visible correction notices to viral deepfakes and push notifications to users who previously viewed or interacted with the content. The implementation of metadata and verified-at-capture frameworks is currently spearheaded in the context of financial and cybersecurity sectors. In particular, the monetary Authority of Singapore (MAS) directives address deepfakes and require financial institutions to use SDKs that verify physical hardware, inspecting camera drivers to ensure video feeds come from real camera sensors (checking for natural optical noise) rather than AI software loops. Singapore is also pushing public and private sectors to adopt C2PA standards for live video, with a movement toward "Verified Capture" using TPM (Trusted Platform Module) to cryptographically sign video streams at the hardware level.

United States: The U.S. has a fragmented approach to synthetic contents, combining federal non-binding guidance, state laws, and industry self-regulation. In 2025, CISA endorsed C2PA Content Credentials as a countermeasure against synthetic media and recommended adoption across government and critical infrastructure media pipelines. California's AI Transparency Act (SB 942, effective August 2026) will require large generative AI platforms exceeding specified traffic thresholds to provide AI-content detection tools and include both visible and latent (invisible) disclosures, subject to implementing regulations. By 2028, the same Act will require manufacturers of capture devices sold in-state to provide hardware-level cryptographic authenticity signals at the moment of capture, supporting later verification of provenance and editing history. However, federal courts have struck down California's mandatory labeling requirements as unconstitutional restrictions on protected expression (e.g., Kohls v. Bonta). In absence of federal mandates, platforms have developed their own policies: Meta bans misleading manipulated media (with parody exceptions) and labels state-controlled media; X requires significantly harmful manipulated media to be removed or labeled; TikTok bans deepfakes of private individuals and requires disclosure of AI-generated realistic scenes; YouTube prohibits deepfakes misleading on political or medical issues. U.S. tech firms are also investing in technical solutions such as Microsoft's Azure Content Authenticity initiative (which builds on C2PA to attach hashes and metadata to content), and Adobe's Content Credentials (which adds provenance info to images). In 2023, a coalition of AI companies pledged to develop watermarking, leading to improvements like OpenAI's invisible embedding.

As a practical baseline for multinational operations, organizations should assume that no single disclosure technique will satisfy every regime. A layered approach - combining visible user-facing disclosure, durable machine-readable provenance metadata, anti-tamper controls, and jurisdiction-specific overlays where exemptions or formatting rules differ - will generally be more resilient than relying on one standard alone.

Kenya, South Africa and UAE: There is no specific requirement for labelling AI-generated deepfakes.

Section **03**

ENFORCEMENT PROVISIONS: ANALYZING RAPID-RESPONSE PROTOCOLS FOR TAKING DOWN HARMFUL SYNTHETIC MEDIA

Brazil: No general legal framework for AI-generated media exists. However, civil liability can arise without a prior court order where a platform fails to act after extrajudicial notification regarding clearly unlawful content. A stronger duty of care applies for serious illegal content. Platforms must maintain accessible reporting channels, consistent moderation policies, transparency reports, and legal representation in Brazil. The current compliance baseline therefore includes: notice-and-takedown workflows for clearly unlawful content (including extrajudicial notifications); high-risk content moderation protocols for serious illegal content and TSE election content; and special handling for paid/boosted and bot-amplified content, given heightened diligence requirements or presumed liability in those cases.

China: Under the Deep Synthesis Provisions and Generative AI Measures, platforms bear strict liability for content moderation. They must have systems in place to identify and block prohibited synthetic content before it spreads, including deepfakes generated without the subject's explicit consent, and any content that subverts state power, endangers national security, or undermines national unity. Platforms that fail to remove non-consensual deepfakes or enforce mandatory AI labeling face severe administrative fines, service suspension, permanent blacklisting from the Chinese internet ecosystem, or immediate criminal charges against executives.

European Union: The DSA imposes comprehensive platform accountability obligations on Very Large Online Platforms ("**VLOPs**"). VLOPs must conduct systemic risk assessments, including on how their services could be manipulated, *e.g.*, through the use of deepfakes to spread illegal content or disrupt civic discourse, and must implement rapid-response notice-and-action mechanisms for illegal synthetic content. The DSA does not itself define what constitutes illegal content; instead, this determination must be made by reference to EU or national laws. Accordingly, synthetic content that breaches the AI Act or other applicable rules may, depending on the nature of the breach and the underlying legal basis, but it should not be assumed that every contribute to a finding of illegality or regulatory non-compliance, failure to satisfy an AI Act disclosure obligation automatically renders the content "illegal content" for DSA purposes. Once content is properly identified as illegal under the applicable framework, VLOPs must act expeditiously upon notice (*i.e.*, they are not required to engage in *general proactive* monitoring). Non-compliance can result in fines of up to 6% of global annual turnover. As of early 2026, the European Commission is actively utilizing these enforcement powers, with formal investigations launched into platforms including X over concerns in relation to reports that its AI chatbot, Grok, may have facilitated the creation or spread of synthetic sexual content.

Looking at Italy as an example of a EU member state approach, we see that Law No. 132/2024 (the national regulatory framework for AI) introduced a new Article 612-quater into the Italian Criminal Code ("Unlawful dissemination of content generated or altered by artificial intelligence systems"), which punishes with imprisonment from 1 to 5 years anyone who causes unjust harm to a person by transferring, publishing, or otherwise disseminating - without their consent - images, videos, or voices that have been falsified or altered through the use of AI systems and are capable of misleading others as to their authenticity. Importantly, where such conduct constitutes a criminal offense, competent enforcement authorities may rely on this provision to pursue the removal of the infringing content from the relevant platform, in addition to any criminal sanctions imposed on the perpetrator. In addition to this new deepfake-specific criminal offense, the applicable legal framework includes

Article 144-bis of the Italian Privacy Code, which specifically addresses the phenomenon of "revenge porn" by allowing individuals who fear the non-consensual dissemination of sexually explicit images or videos depicting them to file a complaint directly with the Italian Data Protection Authority ("Italian DPA"). Upon receiving such a complaint, within 48 hours, the Italian DPA may impose the blocking or removal of the content to digital platforms, by providing them with the relevant order and the reported material or its hash fingerprint. Notably, this mechanism may also be considered applicable to AI-generated or AI-altered intimate content, thereby providing an additional layer of protection against deepfake pornography. Furthermore, the removal of unlawful contents in Italy is also regulated under the notice and action mechanism of the DSA (see above).

India: Under the February 2026 IT Rules Amendments, platforms are strictly prohibited from hosting high-risk "Synthetically Generated Information" (SGI), including AI-generated CSAM, deepfake pornography, and fabricated official documents. Platforms must deploy automated tools to proactively block prohibited content and cannot simply wait for user reports. The "2-hour rule" requires platforms to remove any illegal SGI within 2 hours of notification; failure to comply results in loss of Section 79 safe harbor protection, making the platform - and its local executives - directly liable for the content, with exposure to substantial fines under the Digital Personal Data Protection Act and potential criminal prosecution.

Singapore: As said above, Singapore does not have a single overarching law that addresses synthetic contents. However, under the Online Safety (Relief and Accountability) Act (effective 2026), the Online Safety Commission (OSC) may issue legally binding "Stop Communication Directions" requiring immediate takedown of harmful deepfakes. Under the Protection from Online Falsehoods and Manipulation Act (POFMA), ministers can also force platforms to attach correction notices to viral deepfakes and push notifications to users who viewed the content. Platforms face fines exceeding S$1 million for non-compliance. Chronic failures can result in ISPs being ordered to block access to the platform entirely within Singapore.

United States: There is no overarching federal law mandating general deepfake removal. Instead, the U.S. takes a harm-specific approach. The federal Take It Down Act (2025) specifically targets non-consensual intimate imagery ("NCII"), both authentic and AI-generated, making its distribution a federal crime. Platforms must remove reported NCII within 48 hours or face FTC enforcement actions. At the state level, over 20 states have enacted laws banning deceptive political deepfakes within 60 to 90 days of an election, though several of these bans - including California's - have been struck down by federal courts as unconstitutional restrictions on free speech and protected satire.

Kenya: Enforcement is primarily via the Office of Data Protection Commissioner and there are limited deepfake-specific offences.

South Africa: Cybercrimes Act provides offences, penalties and court orders; enforcement is via the criminal justice system.

UAE: Strong penalties and broad offences under Federal Decree-Law No. 34 of 2021.

In short, for cross-border platforms, the operational baseline should be to triage reports by reference to the shortest potentially applicable removal deadline and the most sensitive content category involved, while preserving a separate legal assessment of whether the content is unlawful in the relevant jurisdiction. Systems designed only around a single notice period or a single harm category may fail where local law imposes faster action, different escalation paths, or regulator-directed corrective measures.

Taken together, this paper shows that regulation of AI-generated content is converging around five core policy questions - training data, creator compensation, transparency obligations, mandatory labelling and enforcement - while remaining fragmented across jurisdictions in legal basis, technical standards, and available remedies. Future analysis should therefore focus on how courts, regulators, and standard-setting bodies address unresolved issues such as lawful access and opt-out mechanisms for AI training, interoperable provenance and labeling standards, cross-border enforcement coordination, and whether further legislative intervention will be needed to harmonize rights, compliance obligations, and remedies in the generative AI ecosystem.

Section **04**

MEDIA LITERACY

Media literacy should be a formal policy pillar in legislation on AI-generated deepfakes, misinformation and disinformation. It is not a substitute for labelling, watermarking, takedown powers or platform duties. UNESCO frames media and information literacy as a national policy issue that should be integrated with education, access to information, freedom of expression, digital technology policy and youth empowerment.

Media literacy helps citizens understand how AI-generated text, images, audio and video can be created, manipulated and distributed at scale. This is especially important because generative AI can make synthetic content more realistic, accessible and persuasive than earlier forms of misinformation. In legislation, media literacy therefore functions as a preventive safeguard: it reduces susceptibility to manipulation before harmful content spreads, rather than relying only on post-hoc takedown, prosecution or fact-checking.

Legal frameworks often require labelling, watermarking or machine-readable provenance for AI-generated content. However, these tools only work if users understand what labels mean, when labels can be missing, and how to verify claims independently. The EU AI Act and Digital Services Act approach deepfakes through transparency, marking and platform-risk mitigation, but expert commentary notes that transparency requirements must be accompanied by user understanding and institutional capacity to interpret synthetic content. Media literacy is also important because misinformation and disinformation legislation can create risks for freedom of expression if enforcement is too broad. OECD stresses that tackling disinformation should not mean controlling information, but should support open, plural and evidence-based information environments.

The countries mentioned in this paper, are all implementing media literacy programmes and some like the EU have also incorporated the need for AI Literacy for the staff of AI developers, AI deployers, AI platforms and AI intermediaries.

Platform providers and AI intermediaries

Platforms and intermediaries should implement user-facing literacy tools because they control the design of sharing, recommendation, labelling and reporting interfaces. OECD highlights the need for platform accountability and transparency as part of information integrity.

They should implement:

- AI-content label explanations;
- friction before resharing suspected manipulated media;
- reporting buttons for synthetic media;
- election-period information panels;
- public transparency reports on deepfake labelling and removals;
- APIs or dashboards for trusted researchers and fact-checkers, where lawful.

AI developers and deployers

AI developers should provide clear user guidance about the capabilities and limitations of generative AI tools, including risks of creating deceptive synthetic media. AI deployers using AI to generate public-facing content should disclose AI use where legally required and should educate audiences on how synthetic content was produced. The EU AI Act's transparency model illustrates this value-chain approach by placing obligations on both providers and deployers for synthetic content and deepfakes.

An overview of the media literacy programmes in the countries are provided below:

- **Brazil:** Media literacy programs are largely decentralized, executed via the National Data Protection Authority (ANPD) and electoral court transparency campaigns designed to battle viral disinformation.

- **United States:** Federal agencies like CISA issue public guides on synthetic media risks, while individual states (eg California and New Jersey) have mandated K-12 media literacy curriculums to teach digital verification.

- **China:** Media literacy in China is state-directed and inextricably tied to internet safety, teaching citizens to consume verified state-media sources and look for mandatory AI labels.

- **EU:** The EU embeds media literacy within the EU AI Act Article 4 and the European Digital Media Observatory (EDMO). It funds member-state educational frameworks focused on critical thinking and recognizing technical provenance markers.

- **Singapore:** Conducted primarily through the Infocomm Media Development Authority's (IMDA) Media Literacy Council, Singapore focuses public education on identifying synthetic media to prevent AI-driven financial fraud and social engineering scams.

- **India:** Managed by the Ministry of Electronics and Information Technology (MeitY), public campaigns and state-led educational programs (such as Kerala's school-level AI training) focus heavily on teaching deepfake verification to rural and non-English speaking populations.

- **Kenya:** Because formal state-mandated digital literacy is still developing, the heavy lifting is handled by civil society collectives (such as KICTANet) and independent fact-checking networks to educate citizens on political deepfakes.

- **South Africa:** Digital and media literacy training is heavily spearheaded by independent watchdogs (e.g., Media Monitoring Africa) and civil organizations, who focus on training voters to recognize deepfakes surrounding democratic elections.

- **UAE:** Driven by the UAE Office for Artificial Intelligence, Digital Economy and Remote Work Applications, the nation prioritizes "AI Literacy" alongside media literacy, training public sector workers and students to use verification tools through initiatives like the UAE AI Camp.

Section **05**

COPYRIGHT & TRAINING: NAVIGATING "FAIR USE" IN THE ERA OF MASSIVE AI DATA SCRAPING

European Union: Unlike U.S. law, EU law does not provide an open-ended "fair use" defense that broadly legitimizes mass scraping of copyrighted content or databases for AI training. The EU framework relies on closed, purpose-bound exceptions, most notably the text-and-data mining ("**TDM**") exception under Directive (EU) 2019/790 (the "**Copyright Directive**"), which permits automated computational analysis of digital text and data to generate information such as patterns, trends, and correlations.

AI training pipelines typically involve reproductions and extractions of copyrighted materials or databases. Whether such acts fall within the TDM exception remains fact-sensitive and legally unsettled, including regarding potential extra-EU application of the Copyright Directive.

The TDM exception may apply only if statutory conditions under the Copyright Directive and relevant EU member state implementing laws are satisfied. The two distinct TDM pathways are briefly outlined below.

	Scientific Research TDM exception (Art. 3)	General TDM exception (Art. 4)
Beneficiaries	Research organizations & cultural heritage institutions	Any entity
Purpose	Scientific research only	Any other purpose, including commercial
Lawful Access	Required	Required
Opt-Out	Not available to rightsholders (mandatory exception) but they may apply proportionate measures to protect the security and integrity of networks/databases.	Available to rightsholders via machine-readable means

The "general" TDM exception under Article 4 is commercially relevant: it permits reproductions and extractions on lawfully accessible content but does not apply where copyright holders have reserved their rights "in an appropriate manner," with machine-readable means for publicly available online content. The Directive does not standardize the technical form of an effective online reservation, leaving AI developers to navigate a patchwork of signals (robots.txt protocols, embedded metadata, contractual terms), none universally recognized as sufficient across the EU.

Several high-uncertainty issues remain unsettled, with outcomes depending on training architecture and access pathways: (i) whether generative AI training qualifies as TDM under the Directive; (ii) how opt-out signaling should be detected through harmonized standards; and (iii) what constitutes "lawful access"—particularly whether breach of website terms or bypassing paywalls negates it (access obtained in breach of terms or by bypassing technical restrictions should be treated as legally high risk).

A risk-mitigating EU compliance approach should involve: (1) implementing robust opt-out detection as a core technical control, including by adopting machine-readable approaches, where suitable also referring to the **GPAI Code of Practice (copyright chapter);** (2) aligning retention and security of TDM reproductions with statutory requirements, as provided under Article 4 of the Copyright Directive; and (3) recognizing AI-output risk even where AI-training is defensible (certain model outputs can independently raise copyright issues where AI-generated content displays substantial parts of protected works).

Regulation (EU) 2024/1689 (the "**AI Act**") does not create a new copyright legal basis for training but imposes obligations on general-purpose AI model providers, including copyright compliance policies and publicly available training-content summaries (Article 53), thereby rendering Copyright Directive compliance auditable. For cross-border models, compliance should be assessed separately for ingestion/access, training/storage/retention practices, and downstream deployment/output distribution—a training posture arguable in one jurisdiction should not be assumed to transfer to another.

For example, under the Italian Legislative Decree No. 177/2021 (amending Law No. 633/1941) there are no substantial additions to the Copyright Directive provisions. Accordingly, the above considerations on EU legal framework apply to Italy as well, just like it is the case for many EU countries.

United States: The United States does not currently have a dedicated federal statute equivalent to the EU AI Act Article 53 that requires AI model providers to publish training-data summaries, implement copyright-compliance policies for training data, or respect machine-readable opt-outs as a specific AI-training obligation. Instead, AI training is addressed mainly through existing U.S. copyright law, especially the doctrines of reproduction rights, derivative works, and fair use, plus ongoing litigation and policy analysis by the U.S. Copyright Office. Unlike jurisdictions that have explicit text-and-data-mining exceptions or AI Act-style transparency duties, the United States relies primarily on judicial interpretation of copyright law.

Singapore: No AI-training-specific copyright mechanism was identified in the current provisions; general copyright law and falsehoods/online harms frameworks remain more relevant.

Brazil, China, India, Kenya, South Africa and UAE: There is no mention in the respective legislation of a U.S.-style open-ended "fair use" defense that clearly and broadly legitimizes mass scraping of copyrighted material for AI training.

Section **06**

CREATOR COMPENSATION: HOW CAN COPYRIGHT HOLDERS EXERCISE THEIR RIGHTS?

European Union: Under the framework described above, compensation demands are likely framed as license-fee or damages claims (in-court or out-of-court) asserting that the TDM exception is unavailable because rights were reserved through opt-out or other TDM conditions were unmet. An AI-training program built solely on a "no opt-out detected" posture faces escalating pressure and risk, especially where AI outputs compete with rightsholders' primary markets. Compensation may be sought through portfolio-level negotiations (collective licensing across multiple works) rather than individual claims; even absent a binding EU-wide compulsory licensing scheme, collective management organizations (CMOs), major publishers, and platform licensors may serve as practical gatekeepers.

The practical question is not only whether remuneration is owed, but **whether licensing, where viable, is the lowest-risk path to reduce uncertainty, maintain lawful EU operations, and avoid material claims** (particularly while opt-out signaling remains fragmented).

The European Parliament resolution of 10 March 2026 on copyright and generative artificial intelligence (2025/2058(INI)) signals continued scrutiny and potential further legislative initiatives on licensing, transparency, and creators' remuneration. Compliance programs should accordingly be designed to withstand both regulatory scrutiny and licensing pressure, and compensation demands should be anticipated as a foreseeable consequence of training governance choices. In this respect, the documentation implementing the transparency obligations under Article 53 of the AI Act should be treated also as documentary evidence possibly used by claimants, regulators, and courts (i.e., to assess adoption of coherent opt-out detection, alignment of training-content descriptions with asserted legal basis, and internal records matching actual practice).

Brazil, United States, China, Kenya, India, Singapore, South Africa and UAE: There is no mention in the respective legislations of an ex ante compensation scheme for copyright holders whose work has been used for AI training. However, in all of these jurisdictions there exists the ex post right for copyright holders to sue for damages and receive compensation for any unauthorized and unlawful use of their works.

ISBN 978-2-9702103-1-3