CREATOR RISK INTELLIGENCE

The Signal

Monthly intelligence on Creator Risk, regulatory change, and brand safety incidents. Issue 20 · August 2025

Meta began rolling out AI content labels on its platforms in 2024. By August 2025 the labels are appearing across Instagram and Facebook on content identified as AI-generated or heavily AI-edited. What the labels show — and what they miss — reveals the limits of platform-level AI detection as a compliance mechanism.

IN THIS ISSUE

01 — PLATFORM AI LABELS: WHAT THEY DO AND DON'T DO

Meta · AI content labelling · August 2025

Meta's AI content labelling system — deployed across Instagram and Facebook through 2024 and 2025 — uses a combination of signals to identify content that has been generated or significantly edited using AI tools. Where the system identifies such content, it applies a label visible to viewers: 'Made with AI' or a similar designation depending on the platform and context.

The system has real value. It provides a baseline disclosure mechanism for AI-generated content at scale — covering millions of posts that no manual review process could reach. For brands running creator programmes, it also provides a partial audit trail: if a creator's content is being systematically labelled as AI-generated and the brand was unaware, that is a gap in the vetting process.

Where platform labels fall short

Detection is incomplete	Meta's system identifies AI-generated content using metadata signals and visual analysis. It does not catch everything. AI tools that strip metadata, use human-in-the-loop editing processes, or operate below the detection threshold can produce content that passes the system without being labelled.

Labels are reactive	The label appears after content is published. It does not prevent non-compliant content from going live or notify brand partners before the content reaches an audience. For a compliance process, reactive detection is a floor, not a ceiling.
Scope is platform-specific	Meta's labels apply to Meta platforms. TikTok has its own labelling approach. YouTube has its own. X has another. A creator running an AI-generated campaign across multiple platforms may be labelled on one and invisible on others. Cross-platform compliance requires a cross-platform assessment.
Brand content is separate	Platform labels apply to organic and creator content. Paid advertising through Meta's ad systems has different disclosure requirements and different detection mechanisms. Brands that think platform labels cover their paid influencer campaign content should verify that assumption.

THE SIGNAL

Platform AI labels are a floor, not a compliance framework. They catch some AI-generated content, on some platforms, after it has been published. A brand compliance process built around relying on platform labels is not a compliance process — it is an expectation that the platform will handle a liability the brand owns.

RECOMMENDED ACTION

Establish your own AI content audit process for creator content before the next campaign cycle. Platform labels are a floor, not a framework. Your brand-level compliance obligations exist independently of what the platform detects.

02 — AUTHENTICITY AT SCALE: THE REMI BADER IMPLICATIONS

Creator authenticity · Brand risk · August 2025

We covered the Remi Bader story in Issue 13 (January 2025) — the body positivity creator who disclosed bariatric surgery after building a significant following on an explicit unaltered acceptance narrative. The August 2025 context for revisiting it is the class action wave that has been building in the US creator economy through the first half of 2025.

US legal analysis firm Morgan Lewis published research in mid-2025 identifying the common themes in creator-related class action filings. Authenticity misrepresentation — where a creator's public personal narrative was materially different from their private reality — features in a significant proportion of cases. The legal theory is straightforward: audiences who engaged with a creator or purchased products endorsed by that creator based on a false authenticity claim were misled.

For brands, the class action risk is indirect but real. A brand that activated a creator specifically because of their authenticity positioning — and where that positioning turned out to be materially false — may find itself named in litigation that connects the creator's misrepresentation to the brand's commercial benefit from the association.

THE SIGNAL

The authenticity claim is the brand asset. When a creator's brand is built on a specific personal narrative, that narrative is what the brand partner is purchasing. If the narrative is false, the brand partner has purchased something that doesn't exist — and may have liability for having presented it to consumers.

RECOMMENDED ACTION

For partnerships where creative rationale is tied to a creator's personal story, build periodic verification into your process. Ask at regular intervals: is the claim that we activated this creator for still demonstrably true? If the answer requires investigation to confirm, that is the gap.

03 — REGULATORY SNAPSHOT

Six markets · August 2025

FRANCE	ARPP H2 2025 focus confirmed: brand-level enforcement actions. Speech-to-text monitoring active. Maximum fine €300,000. Brand liability joint and several. ARPP processing highest volume of enforcement referrals since AI monitoring launched.
GERMANY	UWG private enforcement continues. Competitor brand cases now routine — brands failing disclosure standards face enforcement from peers as well as regulators. Hashtag-cloud inadequate as standalone disclosure. No new legislation, but enforcement is active and accelerating.
ITALY	Ferragni criminal verdict expected imminently — fast-track proceedings extended beyond original timeline. AGCOM Code of Conduct operational. Registration obligations for relevant influencers active. Maximum fine €600,000.
UK	ASA gifting sweep results expected Q3. DMCCA operational. CMA actively monitoring influencer marketing category under new powers. FCA finfluencer enforcement: first formal actions against individual creators and firms expected H2 2025.
EU-WIDE	EU AI Act active. Platform AI labels insufficient for compliance purposes per Commission guidance published August 2025. DSA active. Digital Fairness Act legislative proposal still expected Q3-Q4 2025.
USA	FTC enforcement active across health, supplement, and financial categories. Consumer Reviews Rule operational — fake metrics explicitly prohibited. Class action wave building in creator authenticity misrepresentation category. Morgan Lewis research published mid-2025.

The Commission's August 2025 guidance clarifying that platform AI labels are insufficient for EU AI Act compliance purposes is the most significant regulatory development of the month. It closes a potential gap that some brands had been relying on — the assumption that Meta or TikTok's labelling systems satisfied the Act's disclosure requirements. They do not. Brand-level disclosure obligations are separate from platform-level detection.

STAY AHEAD OF THE SIGNAL

The Signal is published monthly by Signal Intelligence. Each issue covers real Creator Risk incidents, regulatory developments, and the intelligence that brand-side teams need before they need it. Subscribe at signalintelligence.consulting — and if this issue was forwarded to you, make sure you're on the list for the next one.