



Nemko
Digital



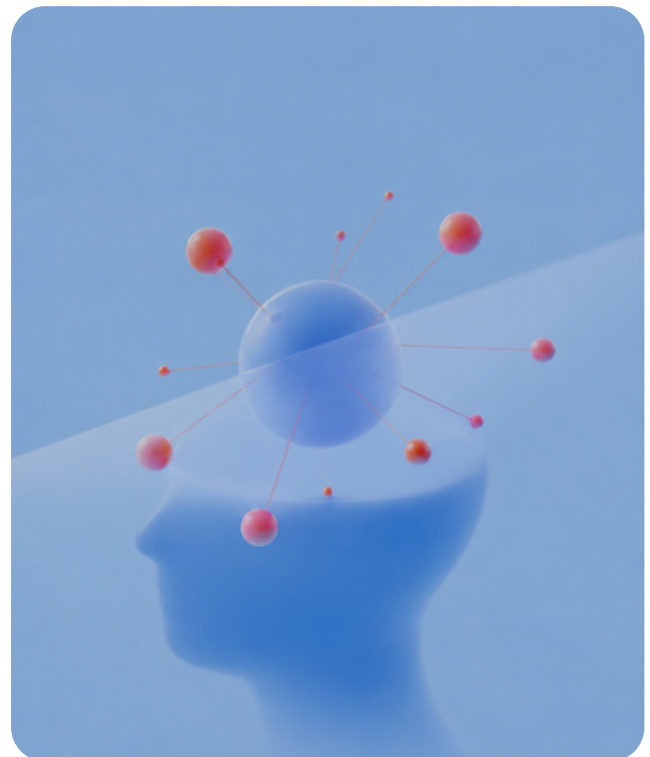
AI Washing in 2026: From Illusion of Autonomy to Proof of Performance

"From the illusion of autonomy to the demand for verifiable performance, AI systems are increasingly defined not by what they claim to do, but by what they can prove."

In recent years, companies have increasingly promoted their products and services as AI-powered, a trend that has given rise to a growing number of so-called AI-washing lawsuits. AI washing can be understood as the strategic overstatement or misrepresentation of artificial intelligence capabilities in products, services, and organisational narratives. At its core, it constructs an illusion of autonomy: outcomes are framed as the result of advanced machine-learning systems, when in practice they may depend on conventional software, manual processes, or extensive human labour.

Autonomy in AI Systems: Levels, Characteristics, and Classification

AI washing occurs when companies use misleading marketing practices to exaggerate the capabilities or sophistication of their technology, thereby creating a false impression of advanced AI-driven innovation. By promoting systems as more advanced or intelligent than they are, companies can create unrealistic expectations among consumers, investors, and regulators. Such practices risk weakening public confidence in AI technologies. Because user perceptions are often shaped by everyday tools such as mobile shopping apps, demographic differences play a significant role in the formation of misplaced trust. Consequently, establishing clear transparency is essential to ensure that consumers can make informed assessments of AI technologies.



The Three Pillars of Machine Autonomy

According to the EU Artificial Intelligence Act, AI systems are designed to operate with varying degrees of autonomy, enabling them to act and make decisions with limited or no direct human supervision. In essence, such systems can perform tasks and navigate complex environments independently, thereby substantially reducing the need for continuous human intervention (see [Recital 12](#)). For regulatory compliance, this operational independence fundamentally shifts corporate liability: businesses can no longer evaluate AI as a static, fully predictable software tool. Instead, [corporate governance frameworks](#) must actively manage the three foundational pillars that define machine autonomy.

The autonomy of such systems typically rests on three foundational characteristics. First, independence of action refers to the system's ability to select among alternative courses of action in pursuit of a goal without being explicitly programmed for every possible scenario. Businesses can no longer claim a system is predictable simply because it is software. Because the AI finds its own pathways to an objective, it introduces unmapped [operational risks](#) that require continuous strategic oversight. Second, operational capability denotes the technical robustness necessary to maintain performance, reliability, and safety standards while functioning independently. Executive boards must treat AI operational health as a core compliance metric. If an autonomous system fails in a real-world scenario, the legal and financial liability rests on the enterprise's leadership, not the machine itself. Third, self-correction captures the system's capacity to detect errors or environmental changes. Adapting its behaviour, accordingly, thereby remaining effective without the need for continual manual adjustment. Managing this "moving target" requires corporate leaders to pivot from traditional, point-in-time software audits to a model of continuous algorithmic assurance.

Levels of Autonomy

The operation of AI systems can be conceptualised along a continuum of autonomy. At the bottom end of this spectrum, systems exhibit low autonomy: the AI functions primarily as an assistive tool that performs narrowly defined tasks upon explicit user request, exemplified by basic spell-checking software. As systems progress toward partial autonomy, they can execute certain functions independently while still requiring a "[human in the loop](#)" for oversight and critical decision-making, as seen in advanced cruise control systems in contemporary vehicles. At the upper end, full autonomy is characterised by the capacity to perceive the environment, process complex data, and complete entire workflows without human input, as in the case of fully self-driving vehicles. The closer AI systems move toward higher levels of autonomy, the easier it becomes for their capabilities to appear almost "magical" to users, particularly when the underlying processes remain invisible or poorly understood.

From Magic to Metrics

The regulatory discourse in 2026 thus converges on a single imperative: AI must transition from a language of magic to one grounded in metrics, evidence, and accountability. This shift has implications for public trust. Repeated experiences with overstated or misleading AI claims can erode confidence not only in individual systems but in the broader AI ecosystem. Conversely, transparent communication about constraints, error rates, and appropriate use cases can foster a more realistic, durable form of trust. This demand for empirical accountability is already being quantified by researchers. For instance, the [2025 Foundation Model Transparency Index \(FMTI\)](#) illustrates this by benchmarking how [transparent](#) major AI developers actually are throughout a model's lifecycle.

FMTI Scores by Release Strategy, 2025

Source: 2025 Foundation Model Transparency Index

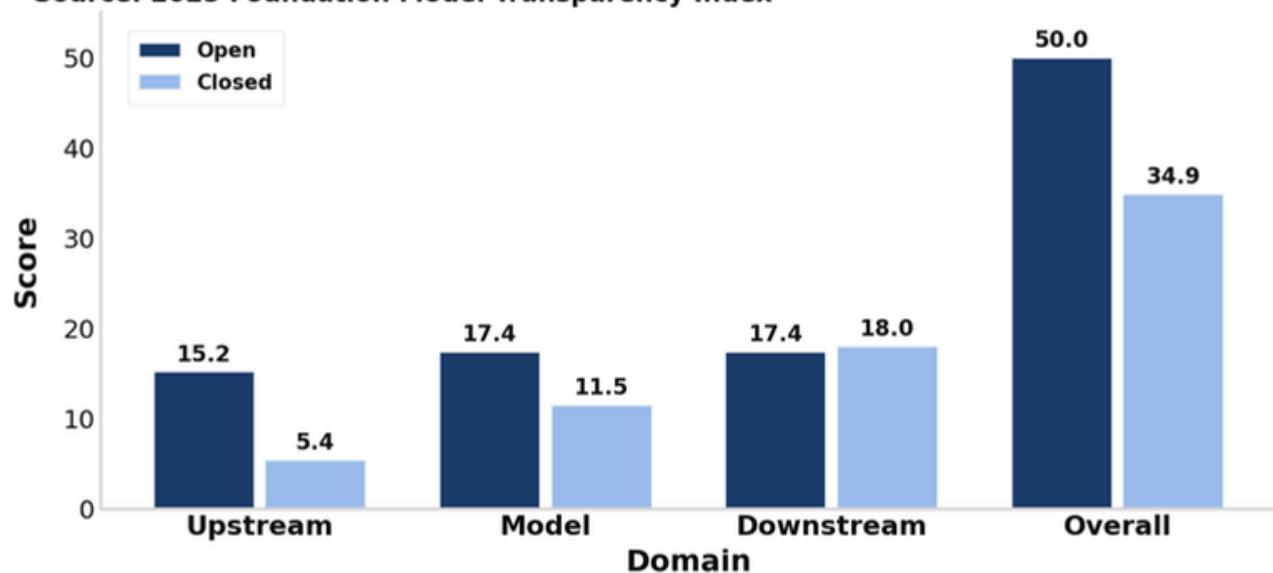


Figure 1: 2025 Foundation Model Transparency Index (FMTI) scores, benchmarking Open vs. Closed AI models across upstream (transparency regarding how the AI was built), model (transparency regarding the model itself), and downstream lifecycles (transparency regarding how the model is used after release).

As regulatory frameworks mature and assurance practices become more standardised, the space for vague claims is narrowing. The trajectory points toward an [AI landscape](#) in which systems are judged less by the promise of automation and more by demonstrated, measurable, and accountable contributions to specific tasks and outcomes.

Unmasking Pseudo-AI

"The AI-washing dynamic echoes the historical "Mechanical Turk", in which a human in fact operated an ostensibly autonomous machine."

The "Human-in-the-Loop" Trap

One of the most frequent forms of AI washing is claiming full automation when in fact only a part of the system is working autonomously. In many cases, companies market products as entirely AI-driven, creating the impression that decisions and outputs are generated independently by AI. However, this representation often obscures the continued and sometimes extensive involvement of human labour within the system. The human-in-the-loop trap is a prime example of this phenomenon. Systems marketed as genuinely AI-driven rely on large, distributed human workforces that review, correct, or complete core tasks. This form of AI washing has significant implications for consumer trust, regulatory oversight, and transparency. It may distort perceptions regarding the true level of autonomy, reliability, and sophistication of AI systems.

Not showing the correct numbers

Companies are facing AI-washing lawsuits not only because the AI does not work as stated, but also because they cannot substantiate the specific percentages used in their ads. This includes selling pseudo-AI, such as a simple computer model, and selling autonomy while providing a sophisticated chatbot, which exemplifies the gap between rhetoric and reality. This exposes providers to severe regulatory risks, including potential prohibitions on offering AI services.

Furthermore, these cases typically do not assert that the underlying technologies are entirely non-functional; rather, they centre on firms' inability to substantiate specific claims. For example, marketing materials and investor communications may highlight the percentage of AI-generated content or the level of automation. From now on, such figures must be supported by a sufficient scientific or empirical basis.

From a business perspective, the inability to substantiate these figures can have significant consequences. Unsupported AI claims may damage investor confidence, undermine customer trust, and expose organisations to reputational harm. In highly competitive markets, being exposed for overstating performance can weaken market position and reduce long-term credibility. Furthermore, regulatory authorities may impose fines or increase scrutiny of future AI products and services.

What are the consequences of AI washing?

Regulators are shifting from a hands-off approach to one that demands hard evidence. Strict AI-washing disclosure obligations are replacing vague, aspirational claims. Companies making AI-related claims to investors and consumers are increasingly expected to substantiate those statements through [assurance and audit mechanisms](#) (insight nemko digital). These assessments must demonstrate that the technology genuinely exists, performs as represented, and has a measurable impact on business operations and performance. This shift draws a clear line between legitimate innovation and deceptive marketing. This broader regulatory tightening is becoming increasingly visible in new legal frameworks that are entering into force.

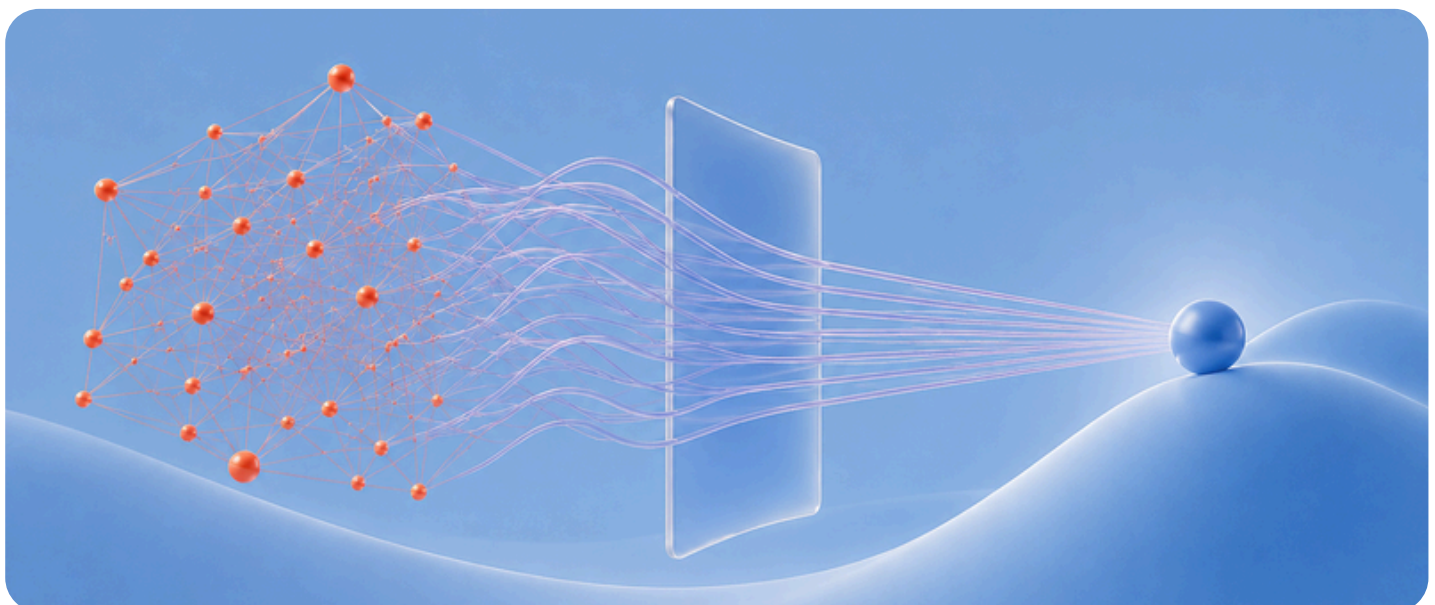
In the United States, state-level initiatives such as the Texas Responsible AI Governance Act and California's SB 53 introduce transparency mandates and substantiation requirements. AI-powered can no longer serve as an empty marketing label; providers must show evidence that AI measurably improves efficiency over standard software.

Internationally, the EU AI Act is widely described as setting a global benchmark against AI-washing. As of August 2026, its full transparency obligations require mandatory labelling of all synthetic content in a machine-readable format as "AI-generated". Misrepresenting a deepfake or AI-written article as "human-made" becomes a direct regulatory violation.

Other jurisdictions are also deploying substantial deterrents against AI-washing. The UK DMCC Act 2024 allows fines of up to 10% of global turnover. AI-related issues are handled indirectly through broader rules on misleading practices. In the United Arab Emirates, initiatives such as Dubai's "AI Seal" create reputational enforcement mechanisms.

Under which non-compliant organisations may lose official certifications. In East Asia, regulators such as the Taiwanese Fair Trade Commission treat deceptive AI-powered claims as false or misleading advertising. Under consumer protection law, with substantial financial penalties and broader regulatory consequences.

In 2026, the central inquiry has evolved from whether a system uses AI to how its capabilities are substantiated, governed, and ethically aligned. The transition from the illusion of autonomy to the proof of performance marks a significant cultural and professional milestone. Today, the integrity of a company's transparency is no longer optional; it must be comprehensive, verifiable, and fair. Truth audits demand disclosure of underlying models, evidence of continuous learning, independent bias and fairness assessments, and rigorous stress-testing of decision logic. Ultimately, these measures are not just about avoiding regulatory fines; they are essential to preventing customer disappointment and mistrust.



The Next Steps to Compliance

To reduce the risk of AI washing and ensure compliance with emerging [regulatory standards](#), organisations should adopt clear internal governance measures regarding the development, classification, and communication of AI systems. Effective governance requires transparency not only in the technical functioning of AI systems, but also in how these systems are presented to consumers, investors, and regulators. The following principles provide a framework for responsible AI-related communication, accountability, and organisational alignment.



Figure 2: A four-step strategic framework for mitigating AI-washing risks, focusing on the systematic transition from marketing claims to verifiable accountability: (1) Classify AI systems, (2) Define AI usage, (3) Support claims with empirical evidence, (4) Align terminology and claims across departments, and (5) Maintain transparency.

1. Establish a clear system classification

Maintain an accurate internal classification of each system, specifying whether functionality is:

- Machine learning-based
- Generative AI-based
- Rule-based automation
- Hybrid or non-AI

All external claims must be grounded in this technical classification.

2. Define and scope AI usage precisely

Ensure that any reference to “AI” in external communications clearly reflects:

- The specific components using AI techniques
- Non-AI logic (rules, templates, human workflows)
- Human-in-the-loop steps

Avoid labelling the entire product “AI-powered” if only one component uses AI.

3. Require evidence for every AI-related claim

All quantitative or qualitative claims relating to AI capability must be supported by:

- Validated testing methodologies
- Reproducible benchmarks or experiments
- Verifiable operational data

Claims without evidentiary backing should not be published.

4. Ensure consistency across organisational functions

Align terminology and claims across:

- Marketing and communications
- Product documentation
- Sales materials
- Investor relations disclosures

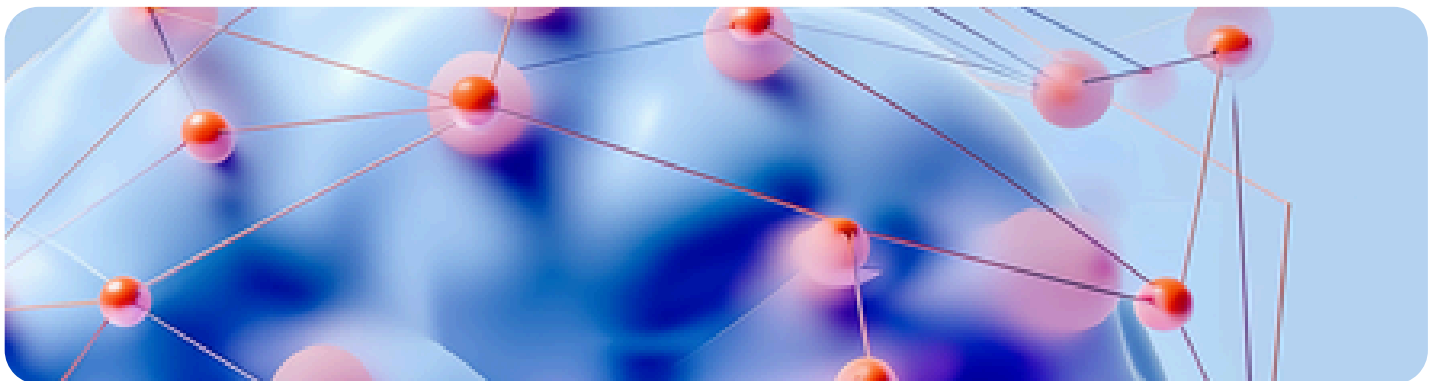
Prevent divergence in the interpretation of AI capabilities across departments.

5. Maintain transparency regarding limitations

Disclose relevant constraints, including:

- Known error rates or uncertainty ranges
- Dependency on human oversight or intervention
- Scope limitations of model performance
- Contextual or domain-specific restrictions

Transparency is a key regulatory expectation in many jurisdictions.



Comparative Table: How Regulators are Spotting "Washing"

Claim Type	Red Flags	What Regulations Focus On
"Advanced / Proprietary AI"	Vague or unclear explanation of what the AI actually does	Whether the claim is truthful and not misleading
Performance claims (e.g., % automation, efficiency gains)	No data, tests, or evidence provided	Whether claims are properly substantiated
"Fully autonomous AI"	Human involvement is hidden or downplayed	Whether the overall impression is misleading
AI- generated outputs	No disclosure that content is AI-generated (where required)	Whether users receive adequate transparency
Ethical / fairness claims	No supporting documentation or testing	Whether governance claims are credible and verifiable

About the author



Joy Haggensburg

Joy is an AI researcher and intern at Nemko Digital. Drawing on her academic research on the EU AI Act and global AI regulations, she focuses on the relationship between AI, trust in technology, and societal impact across different cultural contexts. Her work centers on supporting the safe and responsible adoption of innovative AI technologies across a wide range of sectors, like environmental sustainability, healthcare, and organizational efficiency.