



AI Washing: Need for Better Explainability and Transparency in AI Systems

Author | **Aayushi Gupta**

What is AI Washing?

* AI washing refers to the use of AI-related buzzwords, like “machine learning” or “AI-powered,” to market products that do not meaningfully use AI. Like **greenwashing** or **rainbow washing**, it relies on exaggerated claims to appeal to ethical or progressive values without real substance. *The result: consumers and investors are misled, often paying for innovation that doesn't exist.*



Example: Amazon’s “Just Walk Out” technology markets a seamless experience: shoppers pick items, bypass checkouts, and are automatically billed. However, it is later revealed that this system often requires humans to manually verify a majority of transactions. Is this product actually intelligent? Probably not. Tech companies and products market themselves as using AI, but not doing so, form the basis of ‘AI washing’.

Various forms of AI Washing

1. Automated Misrepresentation: Falsely claiming use of AI for processes that actually heavily rely on human labour.
2. Capability Exaggeration: Overstating AI's true capabilities or failing to demonstrate its specific contribution to an offering.
3. Algorithm Rebranding: Mislabelling traditional automation or rule-based systems as “AI”.

Why is it a problem?

- 1. Many companies take advantage of the public's limited technical understanding. This knowledge gap makes it hard to differentiate genuine AI (like machine learning) from simple rule-based systems marketed as 'intelligent'.
- 2. In India, the absence of enforceable rules, consistent AI definitions, and qualifying criteria makes it easy to misuse the "AI" label.
- 3. This misuse can stifle genuine innovation by masking real breakthroughs behind hype.
- 4. It also erodes public and investor trust, making people more skeptical of true AI advancements.



Greenwashing:

The Guidelines for Prevention and Regulation of Greenwashing or Misleading Environmental Claims, 2024 define greenwashing as a "deceptive or misleading practice, which includes concealing, omitting or hiding relevant information by exaggerating, making vague or false statements, using misleading words, symbols, imagery or any unsubstantiated environmental claims, etc."



Rainbow washing:

"The act of using or adding rainbow colors and/or imagery to advertising, apparel, accessories, landmarks, et cetera, in order to indicate progressive support for LGBTQ equality (and earn consumer credibility) – but with a minimum of effort or pragmatic result."

Source: https://toda.org/assets/files/resources/policy-briefs/t-pb-165_rainbow-washing_oetzel.pdf

Crackdown on AI Washing

INDIAN PERSPECTIVE

India lacks a dedicated law for AI washing, hence, existing regulations against false and misleading advertisements are discussed here.

FRAMEWORK	AUTHORITY	PROVISIONS
Code for Self-Regulation of Advertising Content in India (2024)	Advertising Standards Council of India	<ul style="list-style-type: none"> All descriptions, claims, and comparisons that relate to matters of objectively ascertainable fact should be capable of substantiation Advertisements should not be false, misleading, or deceptive, through exaggeration or implication
Guidelines for Prevention of Misleading Advertisements and Endorsements for Misleading Advertisements (2022)	Central Consumer Protection Authority	<ul style="list-style-type: none"> Establishes clear criteria for permissible advertising (though not AI specifically): <ul style="list-style-type: none"> Should not exaggerate the capability or performance of any good or product Should have a truthful and honest representation Restricts deceptive claims
Guidelines for the Prevention and Regulation of Greenwashing or Misleading Environmental Claims (2024)		<ul style="list-style-type: none"> Environmental claims must be supported by credible evidence, based on independent studies or third-party certifications Mandatory to disclose all material information in the relevant advertisement or communication
Circular on Reporting for Artificial Intelligence (AI) and Machine Learning (ML) applications and systems offered and used by Mutual Funds (2019)	Securities and Exchange Board of India	<ul style="list-style-type: none"> Mandatory reporting requirements with relevant disclosures such as: <ul style="list-style-type: none"> clearly stating if the system utilises AI/ML identifying the specific AI/ML models or tools employed describing the application and its usage of AI/ML description of the application and its usage of AI/ML
Consultation Paper on Proposed amendments with respect to assigning responsibility for the use of AI tools by Market Infrastructure Institutions, Registered Intermediaries and other persons regulated by SEBI (2024)		<ul style="list-style-type: none"> Proposed an amendment to SEBI (Intermediaries) Regulation, 2008, stipulating that any regulated entity employing AI interfaces would bear full accountability for their use, irrespective of the scale of adoption of AI

GLOBAL PERSPECTIVE

Operation AI Comply (2024)

Federal Trade Commission of the United States of America

- Initiative to address misleading and deceptive AI claims
- Took enforcement action against several companies making deceptive claims on the use of AI

CSA Staff Notice 51-365 Continuous Disclosure Review Program Activities for the Fiscal Years (2024)

Canadian Securities Administrators (CSA)

- Disclosures on claims of use of AI systems must be factual and balanced, whether voluntary or mandatory
- Should disclose how it defines AI in its offerings
- No use of overly promotional terms such as "most advanced AI technology"

Automated Vehicles Act (2024)

Parliament of the United Kingdom (UK)

- The Secretary of State can designate specific symbols or words for use with "authorised automated vehicles"
- Prohibits misrepresentation of a vehicle's self-driving capabilities
- Presumes any vehicle not designated as "authorised automated vehicle" as incapable of autonomous driving

AI as a Marketing Term: A quantitative Review of Usage in UK Advertising

Advertising Standards Authority of the UK

- Advises advertisers that when making claims about AI features, they should not mislead consumers or imply capabilities that do not exist
- Advertisers shall ensure responsible advertising

Consumer Protection from Unfair Trading Regulations, 2008 (as amended by the Digital Markets, Competition, and Consumers Act, 2024)

Competition and Markets Authority (CMA) of the UK

- Prohibits false or misleading information or presentation relating to a product that is likely to deceive the consumer
- Empowers CMA to issue enforcement notices and fines up to 10% of global turnover to businesses in breach of consumer protection law, including those practicing AI washing

Case Studies

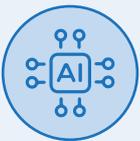
	COMPANY/ PRODUCT	AI INTEGRATION CLAIM	SUSPECTED ACTUAL FUNCTIONALITY	REGULATORY INTERVENTION
False Claim	Delphia Inc. (2024)	<p>Developed algorithms to manage client profiles and provide investment advice</p> <p>AI and ML were used to collect data from clients' social media and online purchases as inputs into these algorithms</p>	<p>Did not collect enough data to provide meaningful insights</p> <p>Did not create any algorithm to use client data to make investment decisions</p> <p>Did not develop the AI/ML capabilities to integrate client data as inputs for algorithms</p> <p>Complete absence of the advanced analytical capabilities advertised as AI-driven</p>	Fined by the United States' Securities and Exchange Commission (SEC)
	Global Predictions (2024)	<p>Investment allocation recommendations through algorithms</p> <p>Incorporated expert AI-driven forecasts</p> <p>First regulated AI-financial advisor</p>	<p>Communicated advice through its chatbot but did not generate advice using AI</p> <p>Did not use AI for forecasts and was unable to provide documentation substantiating AI use</p>	Fined by the SEC
	Rimar Capital USA, Liptz et. al. (2024)	AI-driven automated trading platform for managing client accounts	Falsely described an AI-powered trading system. No real AI system existed	Fined by the SEC
	Kaedim	Proprietary machine learning to convert clients' 2D images into 3D models	Used low-paid human 3D artists to build models, often "fixing up" or entirely replacing the AI output	None reported
	Builder.ai	No-code platform for creation of apps using a virtual assistant	Relied on human work to write code based on customer specifications	None reported
Exaggerated claim	Coca-Cola Y3000	Flavor "created using AI" or "co-created with AI"	<p>Human insights first identified futuristic flavor associations. AI assisted in developing flavor profiles and informed packaging design through AI-generated mood board imagery</p> <p>The AI did not autonomously design the flavor's chemical composition, synthesize ingredients, or conduct taste tests</p> <p>Human scientists carried out the formulation, synthesis, and refinement, drawing on AI analysis as input</p>	None reported
Misleading Claim	Amazon Walk-out Technology	AI-powered cashier-less checkout, using computer vision and sensors to let customers pick items and leave	Could be considered AI-powered if humans only reviewed outcomes, but here, humans performed core tasks that AI was claimed to handle	None reported

Decoding AI

In an era of AI washing, distinguishing true AI from conventional automation is crucial. Addressing it requires demystifying AI-related terms as much as regulating its misuse.



Algorithm- Set of instructions or rules to be followed to perform computer operations



Artificial Intelligence- Artificially simulating human intelligence processes, through computer programs or machines



Autonomous- Use of Technology to perform tasks with minimal human intervention



Chatbot- Program designed to communicate with people through text or voice commands in a way that mimics human-to-human conversation



Human Oversight- Active role of humans in developing, deploying and using AI systems to ensure their ethical, safe, and responsible use



Machine Learning- It is a subfield of AI that enables computer systems to learn and adapt from data- using models and algorithms- without following explicit instructions



Model- Product of AI training, created by running a ML algorithm on training data



Prompts- Instruction entered into a system in a language, images, or code that tells AI what task to perform



* Genuine AI systems are typically grounded in machine learning. They learn from data and adapt their behavior over time, whereas automated systems follow fixed, pre-programmed rules.

Need for Explainability and Transparency of AI Systems

AI washing often capitalises on the public's limited understanding of AI systems. While the implications may be trivial in low-stakes contexts, they become severe in critical sectors such as finance, healthcare, and cybersecurity, where unsubstantiated claims can lead to significant harm. To ensure AI is both effective and trustworthy, two principles are essential: *transparency and explainability*. Transparency refers to clarity into an AI system's architecture, data, and operational processes. Explainability aims to render an AI system's decisions, predictions, or actions comprehensible to humans.

Global frameworks already recognize this need. The OECD AI Principles highlight transparency and responsible disclosure as essential for trustworthy AI, while Recital 27 of the EU AI Act explicitly emphasizes the need for explanations that enable users to interpret and contest decisions made by AI systems. The design of AI explanations typically aims to address challenges associated with the black-box nature of AI, such as distrust, lack of accountability, and limited knowledge discovery, without sacrificing the quality of the AI decisions and actions.

However, achieving effective explainability presents a challenge as to what constitutes a 'good' explanation. Defining a 'good explanation' means moving beyond focusing on technical solutions to meet diverse stakeholder needs— from engineers using explanations to ensure operational reliability, while customers or clients of systems needing it to justify system's outputs. Therefore, explanations demand a whole life-cycle consideration, tailored to relevant actors and specific purposes.

Transparency and explainability standards also require a contextual and risk-based approach. Accordingly, the stringency of these standards must align with the AI system's potential risk level. High-risk applications, such as those in financial forecasting or healthcare diagnostics, thus demand more rigorous standards. In such higher-risk scenarios, requirements should include clear articulation of the system's purpose, data inputs, inherent limitations, and potential biases, alongside explicit user notification of any interaction with or assessment by an AI.

Recommendations

- ✓ Building transparent, verifiable, and explainable AI systems are the key to ensuring safe and ethical development and deployment of AI, delivering the *'whys'* and *'hows'* underlying the AI systems.
- ✓ Fostering AI literacy among all relevant actors is essential. They must be acutely aware of the technology they are dealing with. This awareness includes understanding whether it involves glorified automation or genuine AI.

Achieving such literacy necessitates a dual approach: companies must transparently disclose their AI systems' internal functioning, capabilities, and limitations. Concurrently, users must be empowered to ask the right questions. Such pertinent inquiries include: *What specific problem is the AI intended to solve? At what stages of the process is AI actually involved? Does the system learn and adapt based on new inputs or feedback? And crucially, can the process by which AI reached its decision.*

- ✓ A regulatory framework for AI washing could require a focused and contextual approach rather than an overarching framework, targeting misrepresentation rather than curbing the technology itself.
- ✓ Penalising misleading or exaggerated claims could help curb AI washing.
- ✓ Moreover, independent third-party auditing or verification of such AI systems could serve as a pragmatic safeguard.



The Indian Governance And Policy Project (IGAP) is an emerging think tank focused on driving growth, innovation, and development in India's digital landscape. Specializing in areas like AI, Data Protection, FinTech, and Sustainability, IGAP promotes evidence-based policymaking through interdisciplinary research. By working closely with industry bodies in the digital sector, IGAP provides valuable insights and supports informed decision-making. Core work streams include policy monitoring, knowledge dissemination, capacity development, dialogue and collaboration.

For more details visit: www.igap.in

relations@igap.in | igap.in