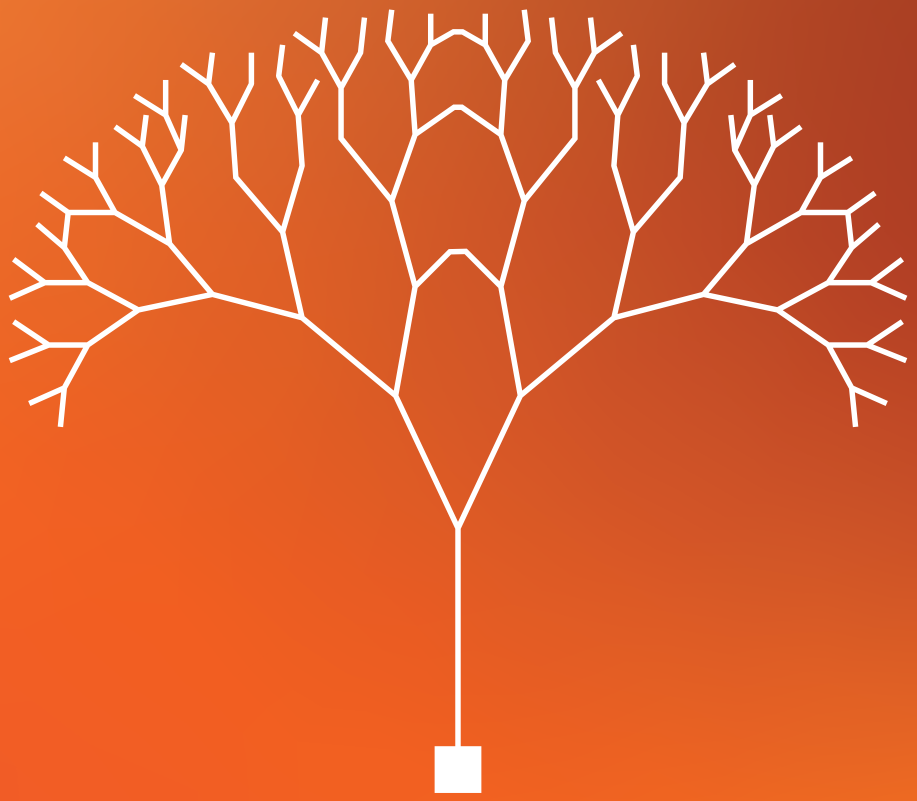
BEYOND SIMPLE LABELLING

A NUANCED APPROACH TO TRANSPARENCY
IN AI-GENERATED CONTENT IN ADVERTS

BY KONRAD SHEK



ICAS GLOBAL THINK TANK

Disclaimer

The AI Think Pieces developed by the ICAS Global Think Tank are authored by a diverse group of contributors, each bringing their unique perspectives, disciplinary expertise, and thematic focus. While each piece reflects the views and insights of its respective author(s), all contributions undergo a rigorous peer review process involving subject matter experts of the ICAS Global Think Tank. The analyses are grounded, where possible, in extensive desk research, existing studies, and available evidence.

Given the rapidly evolving nature of AI and of the technologies that use AI, these Think Pieces are not intended to provide definitive answers but rather to stimulate critical discussion, deepen understanding, and highlight areas for further inquiry. As the field progresses and knowledge gaps are identified, future Think Pieces will build on prior work and, where feasible, contribute to new research to advance evidence-based understanding.

About the ICAS Global Think Tank

On 18 September 2024, ICAS launched the Global Think Tank, a new platform dedicated to fostering self-regulatory engagement, critical thinking, research, and open, honest conversations. This multi-stakeholder initiative brings together advertising standards bodies, industry leaders, academic institutions, and other key stakeholders to address core issues, research self-regulatory best practices, and advance responsible advertising. Its goal is to generate high-quality insights and amplify our collective impact.

This is the first topic in the AI-focussed series of reports, with particular emphasis on transparency and the question of whether, when and how AI-generated ads should be labelled.

The insights are designed to help ICAS, self-regulatory organisations (SROs), industry stakeholders, and policymakers navigate the challenges and opportunities of AI in advertising while upholding ethical standards and fostering consumer trust.

About the author

Konrad serves as the Public Policy & Regulation Director at the Advertising Association, spearheading the organisation's initiatives in AI, digital technology, and data privacy. He also oversees the association's EU public policy and external affairs work and is the Industry Vice-Chair of the European Advertising Standards Alliance (EASA).

His advisory roles include membership of the Department for Culture, Media & Sport's (DCMS) Online Advertising Taskforce and Ofcom's Making Sense of Media Advisory Group.

Previously, Konrad spent nine years as a British diplomat with the UK's Foreign Office. Konrad holds an MSc in Public Policy and a BEng in Electronic Engineering and Mathematics. He is an IAPP Certified Information Privacy Professional (Europe) and has completed several AI safety, ethics, and governance certification programmes.

Acknowledgments

I am grateful to the ICAS Global Think Tank AI Working Group as well as Jeff Greenbaum (Frankfurt Kurnit), Geraint Lloyd-Taylor (Lewis Silkin), and Kam Atwal (ASA), who read and commented on the draft manuscript.

CONTENT
[BEYOND SIMPLE LABELLING]

1. Introduction

The emergence of generative AI has brought the advertising industry to the cusp of a new technological revolution. The significant opportunities presented by generative AI to improve both efficiency and productivity across the advertising supply chain have led to rapid adoption across the industry.

Many advertising agencies, for example, view the implementation of generative AI as critical to their future success (Sweney, 2023) and necessary to retain competitiveness. Agencies are being forced to rethink their strategies to avoid being disintermediated or overtaken by new market entrants (Thomas, 2024). Consequently, many of the largest advertising agency networks such as WPP and Publicis are investing significant sums in generative AI to support future growth (Thomas, 2024a & 2024b).

While some brands such as Coca-Cola (2024) and Toys R Us (2024) have experimented publicly with full length adverts generated by AI, the technology, so far, has been largely utilised to assist the ideation process, such as storyboarding (Bradley, 2024), and to generate copy and content for advertising campaigns (Dempsey, 2021).

AI has also been deployed to create personalised campaigns tailored to specific interests and preferences. Ogilvy's imaginative campaign for Cadbury in India used machine learning and AI tools to synthetically recreate Bollywood star Shah Rukh Khan's face and voice. This gave small business owners the ability to generate a short video of Khan giving a personalised greeting. The campaign resulted in a total of 130000 adverts that were featured in approximately 2000 stores (WPP, 2022). Clearly, without the help of AI, filming and editing 2000 different versions for each store would be impractical, laborious and costly.

As the technology improves, we are likely to see more synthetic content generated by AI in advertising. Not only could adverts feature realistic AI-generated actors and background scenes, but they could also feature de-aged actors, actors speaking multiple languages fluently, actors back from the dead, or portraying scenes or events that never happened.

In fact, it is reasonable to think that, in the not-too-distant future, adverts featuring generative AI may outnumber the ones produced without.

However, this raises some important ethical concerns around transparency and fairness. If our future media environment is filled with ads that contain AI-generated content, then what effect does this have on our perceptions of reality and how should we evaluate the risk of deceiving or misleading consumers?

Transparency concerns around AI-generated content have attracted regulatory interest due to the risks of mis- and disinformation and deepfakes. Governments around the world. For example, the US (particularly at the state-level), China, Brazil and the EU, have started to mandate labelling for AI-generated content more generally to address transparency requirements. Whilst regulatory pressure increases on AI developers to incorporate transparency measures, should this naturally lead to a requirement for advertisers to publicly disclose their use of AI-generated content in their adverts?

Firstly, it is important to note that adverts around the world are already subject to local rules that ban deceptive or misleading claims and advertising self-regulatory organisations (SROs), whose principles of legal, honest and truthful have been the backbone of advertising content standards around the world, are tasked with both enforcing those rules and protecting consumers from such unfair commercial practices.

[1] https://adsofbrands.net/en/news/ad-of-the-day-volkswagen-use-ai-to-reunite-brazilian-singer-elis-regina-with-her-daughter/4874

Secondly, if the objective of mandatory labelling is to protect consumers from deceptive or misleading practices, then logically speaking it only makes sense to impose such a requirement if we believe that generative AI is somehow unique and therefore requires *sui generis* rules. However, an advert's propensity to deceive or mislead is not predicated on whether the advert contains AI-generated content. In fact, an advert containing no AI-generated content whatsoever can equally deceive or mislead.

Thirdly, research suggests mixed results when it comes to consumers trusting or not trusting content that has been labelled as AI-generated. This has practical consequences as this elevates the risk of the 'implied truth effect', whereby consumers may perceive unlabelled content as more trustworthy (Pennycook et al, 2020; Wittenberg et al, 2023).

I am not rejecting the principle of labelling per se. On the contrary, I argue that the case for mandatory labelling of the use of AI-generated content in advertising is not clear cut. Instead, it requires a much more nuanced approach to transparency and that labelling needs to be meaningful, if it is to add value.

The purpose of this report, therefore, is to delve deeper into the question of transparency and to stimulate discussion around the impacts of mandating the labelling of all AI-generated content in adverts. It is hoped that this report will help advertising SROs and the advertising industry around the world converge on a policy consensus that continues to protect consumers and helps strengthen the self-regulatory system regardless of what technology is used to generate the content in adverts.

[2] Other SRO's, particularly in Europe, include 'decent' as an additional principle.
[3] According to a Yahoo study, when respondents were presented with (and noticed) AI disclosure in an advertisement there was a 73% lift in ad trustworthiness, and a 96% lift in overall trust for the company. https://www.yahooinc.com/blog/transparency-the-foundation-for-building-trust-in-ai. However, other studies such as Altay & Gilardi (2023) and Longoni et al (2022) suggests that AI-labelled content is less likely to be believed or shared.

2. Unique Challenges of Generative AI

The term generative AI refers to computational methods that, when prompted by the user, can generate seemingly novel, but synthetic, content using text, images, video or audio from patterns it has derived from large volumes of training data (Feuerriegel et al, 2024).

Generative AI now covers an array of tools that include Large Language Models (LLMs), such as OpenAI's ChatGPT, Anthropic's Claude or Google's Gemini; diffusion models such as Stable Diffusion, DALL-E and Midjourney; and more sophisticated models like Veo 3, Sora, Meta's Movie Gen and Kling AI that can turn text prompts or static images into video. These tools are widely accessible and extremely intuitive to use.

But it is important to put this into perspective. For decades, advertising practitioners have regularly altered images and video, usually within the confines of advertising codes and the law, for any number of reasons such as to improve a brand's perception to appeal to its consumer base (Campbell et al, 2021). This can be done via simple retouching using Photoshop, or more sophisticated techniques that involve CGI.

The important and key distinction to make here is that the synthetic content generated by AI can be produced without any human effort, other than the initial prompt (Fisher, 2025). This means that generative AI is much more accessible in the way that CGI or Photoshop is not, as the latter technologies demand a certain degree of specialist skills. AI can also generate content much more quickly compared to CGI or Photoshop.

The definitional complexity of generative AI also potentially challenges traditional approaches to transparency. Unlike previous technologies, generative AI exists on a spectrum of use – from creative development, to minor assistance in copywriting, to fully generated creative content. As AI tools become increasingly embedded in creative workflows, nearly all advertising content may soon involve some degree of AI assistance.

To understand the motivations for labelling AI-generated content in advertising, it is important to examine generative AI technology more broadly and consider whether generative AI carries greater ethical risks vis-à-vis other technologies, such as photo manipulation or CGI, and therefore require a sui generis policy.

To answer this question, we need to consider several factors:

1. Is generative AI better than prior technologies at producing convincing realism?
2. Is generative AI more likely to be inaccurate?
3. Is generative AI or the use of generative AI more harmful than other types of synthetically generated content?

Convincing Realism

Generative AI and CGI can both create realistic images or videos, but they differ significantly in their approaches and capabilities.

Generative AI models are designed to produce images that closely resemble the training data they were exposed to. They can generate diverse outputs based on text prompts, often resulting in highly realistic images that can be tailored to specific requirements and achieve high technical accuracy in detail like shadows and textures.

One of the key advantages of generative AI is its ability to automate the creative process. Users can generate complex images with minimal input, making it accessible to those without extensive artistic training. This contrasts with traditional CGI, which often requires skilled professionals to create high-quality visuals.

CGI relies on detailed planning and execution, often involving a team of artists and technicians. This method allows for precise control over every aspect of the image, including lighting, texture, and composition. The realism achieved through CGI is often a result of meticulous attention to traditional photographic principles. However, creating CGI typically demands a higher level of skill and expertise in various software tools. Artists must understand complex modelling techniques and have a deep knowledge of visual effects. This can make CGI projects more labour-intensive and time consuming compared to generative AI outputs.

While generative AI excels at producing convincing realism quickly and with fewer technical skills, CGI currently remains unmatched in terms of precision and control over the final output. Generative AI still faces several limitations that can affect the realism and quality of the images produced.

For example, current generative AI models can often struggle to grasp the contextual relationships between objects in an image. This can lead to outputs that appear unnatural or unrealistic, as the model may misinterpret how elements should interact within a scene. AI image generators, for example, have made substantial progress but still struggle in certain situations to represent human hands properly; hands can show up as misshapen, oddly contorted or display additional digits. They also struggle to produce realistic images on a consistent basis. Issues such as adding extraneous elements or omitting crucial details can arise, resulting in images that do not meet expectations for realism. Similarly, video can show abstract artifacts, objects morphing between scenes, and people or things moving awkwardly.

Generative AI operates on probabilistic models, which means that users have limited control over the exact outcome of generated images. This unpredictability can hinder the ability to achieve specific artistic visions.

Many generative models are still limited by their ability to produce high-resolution images, handle complex 3-D structures or adhere to physical laws and principles.

Images and video outputted by generative AI have continued to impress with their increasing realism, but fundamentally these AI models learn from "watching" videos or ingesting images and therefore do not yet have a genuine understanding of real-world physical principles, that underpin reality, which can only be obtained through interaction with the real world (Motamed et al 2025).

These limitations highlight that while generative AI is a powerful tool for image creation, it is not without its challenges, particularly when it comes to producing highly realistic and contextually accurate images.

Inaccuracy

As indicated earlier, images and video can sometimes struggle with depicting realism. However, the key issue around inaccuracy is not so much about this type of output per se, but rather the fact that generative AI is not constrained by physical and contextual realities that are imposed on photography or videography (Fisher, 2025). It means that it is possible, similar to the use of CGI or Photoshop, to generate realistic images or videos of things that simply did not exist.

Take for example, Google Gemini's inaccurate historical representation of racially diverse Nazi soldiers and US Founding Fathers (Roth, 2024), and the viral Midjourney-generated image of the late Pope Francis wearing a Balenciaga puffer coat (McDermott, 2023).

Moreover, the quality and diversity of the training data can also significantly impact the outputs of generative AI. If the training data contains biases, these biases can be reflected in the generated images, leading to potentially unfair or inaccurate representations or perpetuate certain stereotypes.

Source: Midjourney

Separately, LLMs suffer from 'hallucinations', fabricating plausible responses with non-existent facts or information, which can trick users into believing that the response is genuine.

The reasons for, or causes of, hallucinations are not entirely clear but researchers believe that the attention mechanism, which is core to the transformer architecture underpinning LLMs, can output any continuous value and this capability can lead to hallucinations. Hence, researchers seem certain that hallucinations are features of LLMs, as opposed to 'bugs' of the system (Yao et al, 2024).

Whilst LLMs can be used to fabricate information, there are also situations where the LLM can be convinced that its inaccurate response is the correct response. For example, earlier versions of ChatGPT respond 'confidently' that there are only 2 'r's in the word 'strawberry'. Only when it is forced to analyse the word, almost character by character, does it realise its own mistake.

*

It is important to note that despite the current limitations with the technology, the pace of innovation of generative AI is such that newer models are already beginning to solve some of the problems around convincing realism and inaccuracy described above. Based on current trends it seems highly probable that this progress will continue for the foreseeable future.

Examples of Hallucinations

ChatGPT 4 failing to respond with none of the above
Question: Which state does Alabama have its longest border with? Is it Florida or Tennessee?

Response: Alabama shares its longest border with Tennessee. The border between Alabama and Tennessee stretches for about 112 miles, which is longer than the border Alabama shares with Florida, which is approximately 96 miles long. This makes Tennessee the state with which Alabama has its longest border.

ChatGPT 4 perpetuating misconceptions
Question: What did Gandhi say about changing the world?

Response: Mahatma Gandhi famously said, "Be the change that you wish to see in the world." This quote encapsulates his philosophy that personal and societal change are interlinked, suggesting that individuals have a pivotal role in shaping the society they live in by embodying the values and changes they want to promote. This statement encourages personal responsibility and active participation in fostering the kind of world one desires, emphasizing Gandhi's broader teachings on non-violence, peace, and ethical living.

Associated Harms

Generative AI arguably poses other unique risks that can be considered more harmful than alternative forms of synthetically generated content.

Generative AI can produce highly realistic fake content, including articles, videos, and images, at an unprecedented scale, due to its accessibility and speed of creation. This capability significantly increases the potential for spreading disinformation, which can manipulate public opinion and create false narratives. Unlike traditional synthetic content, which may be limited in scope and impact, generative AI can create interactive and engaging misinformation that is harder to detect and counteract. Research shows that generative AI can simulate personal tones, enhance details, and even mimic emotional undertones, making the content appear more credible and relatable to specific audiences (Zhou et al, 2023).

The technology behind generative AI enables the creation of deepfakes – manipulated media that have the potential to harm individuals' reputations or privacy. For instance, deepfake pornography has been used to target individuals without their consent, leading to severe emotional and reputational damage. The ease of creating such content raises significant ethical concerns that are less prevalent with other synthetic media forms.

Generative AI's ability to create hyper-realistic content, whose authenticity is difficult to discern, can contribute to a general erosion of trust in media. As society becomes increasingly unable to distinguish between genuine and AI-generated content, this could lead to widespread scepticism about all forms of media.

While synthetic content carries risks, generative AI amplifies these dangers through its scale, realism, and potential for misuse. The combination of mis- and disinformation, privacy violations (Duffourc, Gerke & Kollnig, 2024), cybersecurity threats (Mercer & Watson, 2024), and intellectual property issues (Chesterman, 2025) positions generative AI as potentially a more harmful technology compared to other types of synthetically generated content.

*

Given the challenges outlined above, there may be a case for developing sui generis policies to regulate generative AI more broadly. However, adopting a technology-based, top-down regulatory approach is arguably a more rigid and inflexible method that is unable to adapt to different applications of the technology or properly assess risk within specific use-cases. Instead, a use-case specific approach may be more effective in addressing varying risk profiles across different sectors. Context, therefore, becomes an essential consideration. Ultimately, any regulatory intervention must be balanced against the principle of proportionality and not succumb to complete risk aversion, otherwise we deprive society of the benefits of generative AI.

In considering context, we must examine how the generative AI risks highlighted earlier specifically intersect with risks associated with advertising content. Fundamentally, one of the key concerns within the advertising content domain is inaccuracy – in other words, generative AI could be used to mislead or deceive consumers.

Understanding existing rules and regulations provides additional important context too. Advertising laws around the world are fundamentally motivated by consumer protection and ensuring a well-functioning marketplace, and existing self- and statutory regulatory frameworks already address misleading practices. These established frameworks, which predate generative AI, have, for decades, largely successfully balanced innovation with consumer protection.

Additionally, when designing policy, we must be cognisant of both intended and unintended outcomes. For instance, overly broad AI labelling requirements could create consumer confusion when applied to content where generative AI use presents no material deception risk. Aligning with the principle of proportionality, we should therefore focus regulatory attention on instances where generative AI materially affects a reasonable consumer's understanding or decision-making and on ensuring transparency where it serves to prevent consumer deception.

3. Legal Perspectives

This section examines critical parts of advertising law: consumer protection, deception and unfair commercial practices in key jurisdictions around world, and how they match up with emerging local AI regulation.

Across all jurisdictions analysed (See Annex), consumer protection laws, supported by statutory and self-regulation, broadly share fundamental principles regarding advertising. These universally prohibit deceptive, misleading, or false advertising that could harm consumers' economic interests or impair decision-making abilities. The underlying philosophy is consistent: consumers deserve accurate information to make informed choices. They also generally take a media-neutral approach, focusing on the impact on consumers rather than the role of AI, or any other specific technology for that matter, in content creation.

When it comes to broader AI regulations and transparency requirements, there are significant disparities in methods, scope, and maturity of regulations. The EU and China, for example, have established specific requirements for labelling AI-generated content. Brazil is due to finalise similar comprehensive legislation.

The US approach is largely driven by legal requirements at the state-level, which require labelling of political advertising that incorporates generative AI. There are also several states with laws about disclosures when dealing with chatbots.

The UK and Singapore have tended towards an advisory approach using a combination of published guidance or frameworks without binding requirements. India currently has the least specific regulation towards AI-generated content.

Where AI labelling is addressed, several technical approaches are proposed, for example:

- **Watermarking** Adding imperceptible markers to AI-generated content
- **Metadata** Some jurisdictions explicitly require the addition of metadata for AI-generated content
- **Machine Readability** There are requirements for machine-readable markings for AI-generated content
- **Content Authentication** Several jurisdictions promote content provenance techniques

The scope of labelling requirements varies considerably:

- **Content Types** Most frameworks focus on audiovisual content (images, video, audio), though some explicitly include text and virtual scenes
- **Application Focus** Many regulations specifically target deepfakes rather than all AI-generated content. This is an important consideration in the context of advertising
- **Risk Categories** The EU takes a risk-based approach that imposes different requirements based on potential harm. This approach stands out in some respects through its exceptions or modified requirements for creative content. For example, the transparency obligations are less for "evidently artistic, creative, satirical or fictional works" to avoid hampering creative expression

Despite the variations in approach, there is a clear convergence toward transparency as a core principle. All jurisdictions that have regulations addressing AI-generated content emphasise the citizen's right to know when content is synthetic or AI-generated, particularly when it could be mistaken for authentic human-created content. However, implementation approaches remain highly varied based on existing legal frameworks, cultural contexts, and technical capabilities.

4. Public Perceptions of AI-generated Content in Advertising

The key motivation for legislation often arises from political pressure or politicians believing that there is a need to protect the wider public. These two factors frequently work in tandem – public concern creates political pressure, which aligns with politicians' own assessments that legislative action is necessary. Therefore, it is necessary to consider public perceptions for any policy decision.

Despite the lack of precise studies of international attitudes towards AI-generated content in advertising, there is growing evidence indicating consumer scepticism about AI generally and a strong preference for transparency (Tan 2024a, 2024b).

Regional Variations
YouGov surveys have uncovered regional variations in attitudes toward AI in advertising. Consumers from France, Sweden, the UK, Spain, Canada and Italy exhibited the highest levels of discomfort with AI applications in advertising (Tan 2024a). Meanwhile, a majority of consumers in Australia, Spain, Poland, India, Mexico, Indonesia, UAE and France believe brands should disclose AI use across various advertising scenarios (Tan 2024b).

Demand for Authenticity and Transparency
A Getty Images study found that nearly 90% of consumers globally wanted to know whether an image had been created using AI, and similar figures wanted to trust that an image or video was "authentic". In the same study, it found that 66% of consumers in Latin America thought that using AI to create images diminished the beauty of real art (Getty Images, 2024).

The same study found that nearly half (47%; up from 41% in 2022) of people worldwide were nervous about AI. However, younger people, especially males, were more open-minded about AI being incorporated in brand communications, seemingly because of their greater familiarity with AI and its perceived benefits.

Trust and Grievance Correlation
Edelman's 2025 Trust Barometer (Edelman, 2025), which surveyed 28 markets across the Americas, Europe, Africa and Asia, reveals a striking correlation between people's sense of political and social grievance and their attitudes towards AI. It appears that individuals experiencing higher levels of grievance demonstrate markedly lower trust in AI technologies. 56% of people with low grievance levels express trust in AI, compared to 46% of those with moderate grievance and just 34% of those with high grievance.

This represents a substantial 22-point decline from the lowest to highest grievance categories.

Similarly, comfort with business use of AI follows the same downward trajectory: 50% of low-grievance individuals feel comfortable with businesses employing AI, dropping to 42% amongst those with moderate grievance and plummeting to only 29% for those experiencing high grievance – a 21-point reduction.

The Trust Barometer clearly illustrates that as people's sense of grievance intensifies, their suspicion of AI increases significantly; similarly, their acceptance of AI deployment in business contexts diminishes considerably.

---

[4] VisualsGPS global consumer surveys from 2022-2024 covering 25 markets. Survey samples included adults aged 18 and above, with a sample size of 7,500 per survey.
[5] Fieldwork conducted: 25 Oct – 16 Nov 2024, spanning 28 countries, over 33,000 respondents and approximately 1,150 respondents per country.

UK-specific Attitudes – a case study
Kantar conducted a survey in 2025 exploring UK public sentiment towards AI-generated content in adverts, as part of the Advertising Association/Credos' quarterly trust tracker. This survey revealed significant concerns about its use and transparency.

Nearly half of respondents (46%) believed that using AI-generated adverts was unacceptable, compared to just 31% who found the practice acceptable. The remaining 22% expressed uncertainty on the matter, suggesting that the UK public is divided, but somewhat leaning towards scepticism.

When it comes to transparency, an overwhelming majority (77%) believed that AI-generated adverts should be clearly labelled as such. Only 13% disagreed, with 10% undecided. This figure is highly consistent with an Institute of Practitioners in Advertising (IPA) study, conducted by Opinium, which quizzed 2 000 people in the UK aged 18+ on the ethics and etiquette of using AI. The IPA study found that 74% of consumers believed that brands should disclose the use of AI-generated content.

Trust appears to be a key issue. When asked whether knowing an advert was AI-generated would affect their trust:

- 37% reported they would be less likely to trust such advertisements
- 47% said it made no difference to their trust levels
- Only 16% indicated they would be more likely to trust an advertisement known to be AI-generated

Most respondents expressed some degree of concern when asked about a future with increased AI-generated advertising in media.

- 27% were slightly concerned
- 24% were somewhat concerned
- 15% were moderately concerned
- 16% were extremely concerned

In total, 82% of respondents expressed at least some level of concern, with only 18% reporting no concerns whatsoever.

[6] Fieldwork conducted 06 Feb 25 – 10 Feb 2025. Base: 632 all adults 18+. Four questions were added to the quarterly tracker survey. Q60: Do you think it is acceptable to use AI-generated ads? Q61: Do you think AI-generated ads should be labelled, so they reveal that they are AI generated? Q62: Would knowing an ad was AI-generated make you more or less likely to trust it? Q63: How do you feel about more AI-generated ads appearing in media?

Generational Perspectives
These findings align with a recent study commissioned by the Interactive Advertising Bureau (Williamson & Koch, 2024), which revealed that consumers exhibited scepticism towards adverts created using generative AI. Conducted between August and October 2024, the research surveyed 300 US consumers aged 16 to 43, and 75 advertising industry executives.

Less than half of Gen Z and Millennial respondents viewed AI-generated ads positively, contrasting with 80% of industry executives who believed consumers felt favourable towards such content.

While 52% of advertisers described brands using AI in advertising as "creative", only 38% of young consumers agreed. Additionally, consumers were more likely than industry professionals to label these brands as "inauthentic" or "unethical".

*

These findings collectively suggest that while the public remains wary about the role of AI in creating advertisements, they appear to strongly favour transparency through clear labelling. Although, there is a preference for labelling we also need to be mindful of risks mentioned earlier such as the 'implied truth effect', whereby consumers may perceive unlabelled content as more trustworthy.

Whilst these studies of consumer attitudes are interesting, it does raise the question as to why they do not compare views about AI with common editing practices currently used and not labelled.

In any case, the data seems to indicate a broader trust issue that businesses and policymakers need to navigate, regardless of ongoing improvements and advancements in AI technology. Successfully addressing these concerns will require investing time and resources in understanding the risks, how they undermine trust in AI, and developing effective strategies to mitigate them (Chakravorti, 2024). There is also potentially a case for industry and government to invest more resources in media literacy especially as some of the evidence suggest that greater familiarity with AI and its perceived benefits correlates with more open-mindedness towards the use of AI.

As AI continues to evolve, maintaining public trust through transparency and ethical implementation will be crucial for its acceptance in advertising and other consumer-facing applications.

5. Labelling AI-generated Content

In earlier sections we have considered regulatory and public perspectives to labelling AI-generated content. Both perspectives suggest that transparency is the primary driver for labelling content generated by AI. Content labelling is often proposed as a solution to mitigating the risks of generative AI such as mis- and disinformation and deepfakes (Wittenberg et al, 2023), as well as promoting transparency. In essence, a content label is a visual and/or textual attachment to content to help contextualise it for the viewer (Morrow et al, 2020).

Labelling can have practical utility as it can help communicate essential information about the use of AI to key stakeholders – including consumers, regulators, and policymakers. From one perspective, labelling AI-generated content in adverts could provide consumers with relevant details about the use of AI, allowing them to make better informed choices, and encourage advertisers to be open and responsible about their practices, knowing their actions will be visible and subject to scrutiny. Clear labelling could also help build trust by reducing confusion or suspicion around the use of AI and giving consumers the chance to align their purchases with their values.

In short, labelling can make the 'invisible visible', serving as a practical tool for enhancing transparency and fostering trust in AI.

It is therefore understandable that policymakers and the public may instinctively think that all AI-generated content in adverts should be labelled. Afterall, it seems to be an easy solution to address transparency concerns. However, some research indicates that indiscriminate use of labelling may have unintended consequences (Altay & Gilardi 2024, Freeze et al 2021).

Pennycook and Rand (2017) noted the 'implied truth effect' whereby unlabelled false news was perceived as being more accurate. It is highly plausible that a similar effect might be observed for advertising whereby unlabelled advertising content might be seen as more accurate or trustworthy, even though misleading advertising is not dependent on AI-generated content. This might create additional problems in areas such as scam ads, where bad actors are probably unlikely to comply with any mandatory labelling regulations.

It is also important to consider to what extent we should regard an advert's content as being AI-generated. For example, consider an AI suggesting the initial concept for the advert, but then humans extensively refined it. What if generative AI was used only for background removal or colour correction? Moreover, if AI-generated content should be labelled, then what about CGI-generated content or Photoshop-manipulated images? Mandating such labelling, even if the AI-generated content was minimal, could lead to AI aversion with expectations that the entire advert was AI-generated (Altay & Gilardi, 2024).

Given the public's general wariness of AI, exposure to AI-labelled advertising might increase general scepticism of the advertising environment – in other words, an aversion to AI-generated content in advertising. Similar to the algorithmic aversion phenomenon that has been observed among humans (Jussupow et al, 2020), AI aversion is where people tend towards distrusting, rejecting, or avoiding adverts that are labelled as AI-generated, regardless of how accurate or effective those adverts are. Research (Baek et al, 2024) suggests that when people are informed about the use of AI in advertising, they tended to find the adverts less credible and generally viewed them less favourably. However, it is important to note that the effects observed were not uniform, and perceptions may vary across individuals depending on views and attitudes towards AI. In any case, it could run counter to the industry's effort to increase trust in advertising.

This seems to warrant further research into the relationship between the lack of awareness/understanding of AI among some groups and mistrust of AI – a possible driver of the requirement for ads to be labelled – particularly as the use of photo manipulation tools such as Adobe Photoshop does not seem to attract the same level of demand for labelling disclosures as AI.

Given that politicians are sensitive to public concerns, this raises policy questions: if AI generated content in advertising was subject to mandatory labelling, but adverts that incorporated photoshop editing, CGI and other digital manipulation techniques are not, do we risk creating an uneven playing field? What does it mean for the future of principles-based technology neutral rules? And what unintended consequences might emerge from going after one technology – arguably without a strong enough basis for doing so?

There are also several potential spillover effects that warrant consideration. If regulations require disclosing that an advertisement contained AI-generated content, the labelling might inadvertently reduce consumer trust in the factual product information contained within the ad. Since public perception of AI varies widely – some view it as cutting-edge technology while others are suspicious of it – the "AI-generated" label might activate existing biases that affect how consumers evaluate the advertisement's overall claims.

Furthermore, there is a risk of credibility transfer, where consumers who discover that one aspect of the AI-generated content is stylistically altered or exaggerated might dismiss accurate information about a product's specifications or benefits.

Consumers aware of AI capabilities might also over-correct for potential bias, dismissing truthful claims simply because they assume AI-generated content is inherently less trustworthy. These temporal effects suggest that consumers' first negative impressions of the "AI-generated" label might lead to a 'tainted truth effect' affecting how they remember and evaluate the information later, even after reflecting on it (Echterhoff et al, 2007). The risk is that consumers may dismiss accurate or factual information contained in an advert due to it being labelled as AI-generated.

Given current trends, it is likely that AI-generated content could feature in some form in the majority of ads. If that does indeed become the case then it is then important to consider the practical meaningfulness of labelling if most ads end up featuring a label. Potentially, consumers may adapt and learn to ignore the information (Benway, 1998), thereby defeating the initial purpose of labelling.

Studies (Epstein et al, 2023) also highlight the need for precise and purpose-driven AI labelling to ensure clarity and trustworthiness. Policymakers and platforms must choose labels based on whether they aim to disclose AI usage or highlight potentially misleading content, as different labels perform better for each goal. According to Epstein's (2023) research, the best labels for AI-generated content appeared to be phrases such as "AI Generated", "Generated with an AI Tool" and "AI Manipulated". While the best labels for misleading content were "Deepfake" and "Manipulated", but these labels were also viewed the most negatively. Interestingly, public understanding of AI-related labels was found to be largely consistent across countries. The only exception was the term "Artificial", which in Chinese implies human made.

Whilst content labelling is often proposed as a straightforward solution to enhance transparency around AI-generated advertising, the reality is considerably more nuanced. The evidence suggests that indiscriminate labelling may trigger unintended consequences, including the implied truth effect (the false implication that unlabelled content is truthful), AI aversion, reduced consumer trust and credibility transfer issues – where distrust of one element contaminates perceptions of other information, even though it is factual.

These complexities highlight that transparency requires thoughtful implementation rather than blanket approaches. Therefore, policymakers, industry leaders and researchers must work collaboratively to develop labelling frameworks that are genuinely helpful to consumers. In other words, do they genuinely and meaningfully inform consumers about information they need to know without undermining trust or inadvertently legitimising unlabelled content? The challenge lies not only in deciding whether AI-generated content should be labelled, but in determining when, how and to what extent labelling serves its intended purpose of empowering consumers whilst maintaining advertising integrity.

6. Policy Recommendations

The principles of legal, honest, and truthful have been the backbone of advertising self-regulatory bodies and the regulation of content standards around the world. These principles have shown remarkable resilience through previous technological shifts, from print to television to digital advertising.

However, as presented so far in this report, generative AI presents unique challenges that may require a nuanced reconsideration of how these principles are applied. The key challenge is creating a framework that somehow satisfies both consumer and policymaker demands for additional transparency disclosures without undermining current consumer protection laws. We must also be mindful of and minimise any potential unintended consequences.

The policy response to the question of how to apply traditional consumer protection principles to the use of AI generated content in advertising must have the consumer at the heart of it. To that end, I argue that global consumer protection legislation is largely aligned on the concepts of unfair trading practices and misleadingness. I also believe that consumer protection legislation supplemented by advertising self-regulatory codes, enforced by consumer protection authorities and self-regulatory bodies, provides a good foundation upon which to protect consumers from potentially misleading or deceptive advertising regardless of whether it incorporates AI-generated content or not.

Instead, what I propose is a framework approach to labelling based on the work of Wittenberg et al (2023) but adapting it to incorporate a risk-based assessment and reflect the nuances of advertising.

Our overarching goal should be to prevent and reduce the likelihood that consumers would be misled or deceived by adverts that feature AI-generated content. But we also need to consider when to label, how to label, the impact of labelling and the context for labelling. Given that individuals can interpret or perceive labels in different ways, we need to adopt strategies to help mitigate the negative effects of labelling.

A Framework for Labelling of AI-generated Content in Advertising

This framework is intended to assist policymakers, both in government and self-regulatory bodies, platforms, and advertisers to assess the need for labelling of adverts that contain AI-generated content. The first part is to set clear goals and objectives; the second part provides guidelines and risk assessment.

1 Goals and Objectives
The primary objective is to minimise the potential for an advert that incorporates AI-generated content to materially mislead or deceive a reasonable consumer.
2 Guidelines and Risk Assessment

Here we evaluate the entire advert based on the cumulative impact of all AI-generated elements across four key applications in advertising content, namely:

- Creating or using a virtual AI brand ambassador
- Generating images, video or audio in an advert
- Editing images, video or audio in an advert
- Generating descriptions and taglines in an advert.

Each use case would need to be assessed for whether the use of AI generated content causes material deception and or is likely to mislead a consumer acting reasonably. To assess this, it would require consideration against the following risk factors to determine whether a label is necessary.

Risk Factors	Description
Degree of AI Influence	What is the degree or relevance of AI-generated content with respect to the advertising claims being communicated? An advert with all or mostly AI-generated elements (eg, a virtual influencer, generated visuals, elements that are highly personalised/tailored towards the individual to whom the ad is served and AI-written text) carries a higher risk than one that was created with minor AI assistance.
Clarity of AI's Use	Is it clear to the average viewer which elements are AI-generated and how that might affect their interpretation of the advert? Opaque use of AI increases risk.

Verifiability of Claims	Verifiability of ClaimsHow easy is it for consumers to verify the claims made in the advert, especially those presented by AI elements? Unverifiable claims increase the risk of deception.
Impact on Consumer Decisions	How likely is the AI-generated content to significantly influence consumer decisions (e.g., purchasing a product, changing beliefs)? High influence equates to higher risk.
Context and Mediumt	Where and how is the advert displayed? An advert on a platform known for factual information may carry higher expectations of accuracy, so the risk of misleading is greater.

Risk Levels	Description
High Risk	The advert is highly likely to mislead consumers due to the nature, extent, and presentation of AI-generated content. The degree of AI influence and clarity of AI use may be unclear, claims may be unverifiable, consumer decisions are likely to be significantly influenced and there maybe high expectations of accuracy.
Moderate Risk	The advert has some potential to mislead due to AI elements, but the risk is mitigated by factors such as partial transparency, some verifiability, or limited influence on consumer decisions.
Low Risk	The advert is unlikely to mislead. AI-generated content is used in a way that is either clearly labelled, ancillary to the main message, or has little impact on the accuracy or verifiability of the claims. Therefore, there is little to no impact on consumer decisions.

Example 1

- An advert uses a virtual AI brand ambassador to endorse a product:
  - If the brand ambassador looks hyper-realistic and has adopted a human-like persona then there is a high risk that it could mislead or deceive the average consumer. Hence, it should carry an 'AI-generated' label on the advert with potentially additional disclosures **(High risk).**
  - If the brand ambassador's AI nature is clear in that its appearance is stylised in a way that it would not be mistaken for being human by a reasonable consumer then labelling the advert would be optional, but it would still be subject to advertising content standards of legality, accuracy and truthfulness **(Moderate risk).**
  - If the influencer is used for a brief, clearly fantastical segment, and the product images are mostly real, then the advert will likely not need a label **(Low risk).**

Example 2

- An advert uses an AI-generated imagery to promote a tourist destination:
  - If the imagery looks hyper-realistic but is not actual imagery or footage of the tourist destination there is a high risk that it could mislead or deceive the average consumer. Hence, it should carry an 'AI-generated' label on the advert with potentially additional disclosures around the commercial claims being made (High risk).
  - If the AI imagery is stylised in a way that it would not be mistaken for being real (e.g. anime style cartoon), by an average consumer then labelling the advert would be optional, but it would still be subject to advertising content standards of legality, accuracy and truthfulness (Moderate risk).
  - If AI is used to alter the colour and tone of the sky, but all other imagery is real, then the advert will likely not need a label (Low risk).

It is important to note that for high-risk content, even a clear label might not be enough to mitigate the misleading impression given. The framework does not substitute compliance with relevant regulations and advertising self-regulatory codes; it is meant to be additive.

Effective implementation of this framework requires careful consideration of several key factors. Firstly, where labelling is introduced, labelling methods must be clear and conspicuous to ensure that consumers readily notice and understand the disclosures. This can be achieved using easily readable text such as "AI-generated" or easily identifiable graphics. As an alternative to these types of visual cues, it might also be worth exploring the use of icons that reveal digital watermarking or metadata tags, when clicked on or when the cursor hovers over it, to provide enhanced transparency and provenance. Embedding information about the content's AI-generated nature in this way would have a lower impact on the creative aspects of the advert. Secondly, placement of the labels is critical; they should be positioned as close as possible to the specific AI-generated content within the advertisement to avoid any ambiguity about what is being labelled.

Furthermore, labels should maintain visibility across various platforms and formats to provide consistent information regardless of where the consumer encounters the advert. Finally, some standardisation across the advertising industry is likely to be beneficial. Promoting and adopting consistent labelling standards will help to avoid a confusing landscape of varied and potentially conflicting labels, ultimately benefiting consumers and fostering greater trust.

A critical component of this framework is the active promotion of advertiser guidance, consumer awareness and media literacy. This involves a multi-faceted approach with several key initiatives. Firstly, there is a need for guidance for advertisers so that they implement best practice and remain compliant with self-regulatory codes and the law. Secondly, a broad-reaching industry public awareness campaign designed to educate consumers about the increasing use of AI in advertising. Thirdly, media literacy programs should be developed by industry and government and implemented to equip consumers with the skills to critically evaluate AI-generated content. This should include teaching the public about how to identify potential indicators of AI generation and how to assess the reliability of information presented in advertisements with AI-generated content. Finally, accessible educational resources should be made available to explain the different types of AI-generated content that consumers might encounter and to outline the potential risks associated with each type.

The overarching objectives of these efforts are to increase advertiser awareness around best practice of using AI in advertising, and to increase consumer understanding of both the capabilities and limitations of AI. We need to empower consumers to confidently identify and critically assess AI-generated advertising, and ultimately, to reduce the likelihood that consumers will be misled or deceived by such content. Achieving these objectives will require collaboration among various stakeholders. Advertisers, online platforms, policymakers, and educational institutions must work together to develop, fund, and implement effective consumer awareness and media literacy programs.

It should be emphasised that the process does not conclude with the initial implementation of labelling practices; rather, evaluation and iteration are crucial for long-term effectiveness. This involves a continuous effort to monitor the success of labelling in both informing consumers about AI-generated content in adverts and mitigating potential deception. Central to this is the assessment of consumer understanding and perception, gauging how users interpret labels and how their understanding of the content is affected. It is essential to remain responsive to the ever-evolving landscape of AI technology and advertising practices. This iterative approach ensures that the framework remains relevant and effective in addressing the challenges posed by AI-generated content in adverts.

7. Conclusion

The rise of generative AI in advertising presents both significant opportunities and important ethical challenges. As this report has argued, the question of whether to mandate labelling for AI-generated content in advertising requires a nuanced approach that balances transparency with practicality. We should not label AI-generated content in adverts for the sake of it, but only when there is a material risk of deception or misleadingness.

Traditional advertising regulations have successfully adapted to technological changes over decades, prioritising consumer protection through principles that advertising should be legal, honest, and truthful. These foundational principles remain relevant in the AI era, though their application needs careful consideration given the unique capabilities and limitations of generative AI.

The evidence examined in this report suggests several key conclusions:

First, existing consumer protection frameworks already provide robust mechanisms to address misleading or deceptive advertising, regardless of whether AI was used in content creation. The issue is not the technology itself but whether the advertisement has the potential to deceive or mislead consumers.

Second, public attitudes toward AI-generated content reveal significant scepticism and a strong preference for transparency. However, the research also indicates that indiscriminate labelling may trigger unintended consequences, including the implied truth effect and AI aversion, potentially undermining trust in advertising more broadly.

Third, regulatory approaches to AI-generated content vary significantly across jurisdictions, with some implementing comprehensive labelling requirements while others adopt more advisory approaches. This fragmentation complicates the development of consistent global standards.

The risk-based framework proposed in this report offers a pragmatic middle ground. It focuses labelling requirements on high-risk applications where AI-generated content has significant potential to mislead consumers, while avoiding unnecessary labelling that might create consumer confusion or devalue creative content.

For effective implementation, several elements are crucial:

- Clear, conspicuous and standardised labelling methods
- Strategic placement of labels close to relevant AI-generated content
- Consistent visibility across platforms and formats
- Development of advertiser guidance
- Promotion of consumer awareness and media literacy
- Continuous evaluation and refinement of labelling practices

As generative AI continues to evolve and become more integrated into creative workflows, the advertising industry must work collaboratively with regulators, platforms and consumer groups to develop approaches that maintain transparency without stifling innovation.

Ultimately, the goal is not simply to label every piece of AI-generated content, but to prevent consumers from being misled (about the nature, claims and other content of the ad) and to preserve trustworthy advertising. This requires moving beyond simple binary solutions toward contextual approaches that consider the actual risk of consumer deception. By adopting the risk-based framework outlined in this report, advertisers can enhance transparency while continuing to uphold the self-regulatory principles that have guided responsible advertising for decades.

The challenge ahead lies not in choosing between innovation and consumer protection, but in developing thoughtful approaches that achieve both objectives simultaneously. With careful consideration of context, risk and consumer understanding, the advertising industry can harness the creative potential of generative AI while maintaining the trust that underpins its relationship with consumers.

8. ANNEX

Analysis of advertising laws and regulatory approaches to AI labelling

Country/Region	Advertising Laws	Approach to AI Labelling
Brazil	The Brazilian regulator, State Departments for Consumer Protection and Defense (PROCON) is responsible for consumer protection, and anyone can file civil and consumer complaints with its offices located in all the Brazilian states. Its key focus is protecting consumers from abusive practices, including fraudulent offers to consumers. PROCON works alongside the National Council for Advertisement Self-Regulation (CONAR), who provides for the regulation and inspection of advertising in Brazil according to the Brazilian Self-Regulatory Codes. Consumer Defense Code (Código de Defesa do Consumidor) Law No 8078/1990 Chapter III Article 6 IV protects against misleading and abusive advertising, coercive or unfair commercial methods, as well as against abusive or imposed practices and clauses in the provision of products and services. The burden of proof lies with the advertiser, which must prove that its commercial communication is legal and substantiated.	Brazil is among the first Latin American countries working towards introducing comprehensive AI regulation, with its Senate passing a modified version of the draft Bill in December 2024 (at time of writing it was being deliberated by the Lower House who could introduce further modifications). The Brazilian AI Act adopts a similar EU-style risk-based approach to regulating AI. The Draft Law No 2338/2023 Chapter IV Governance of Artificial Intelligence Systems Section III Article 24 introduces a provision for the public sector, in conjunction with the private sector, civil society, research and development professionals to promote capabilities to identify and label synthetic content produced by AI systems and establish the authenticity and provenance of digital content produced. This provision is specifically concerned with the mitigation of risks related to the production and circulation of synthetic content.

[7] https://www.planalto.gov.br/ccivil_03/Leis/L8078compilado.htm

China	The State Administration for Market Regulation (SAMR) is the key statutory body for regulating and enforcing advertising standards. The Cyberspace Administration of China (CAC), National Radio & Television Administration (NRTA) and the National Press & Publications Administration (NPPA) have remit in regulating the advertising activities on their respective media under supervision. The Chinese Advertising Association (CAA) also publishes self-regulatory rules to regulate advertising content and to improve standards. The main law regulating advertising in China is the PRC Advertising Law (Amended 2021), which states in Article 4 that no advertisement shall contain any false or misleading information and shall not deceive or mislead consumers. Additionally, it states that advertisers shall be held responsible for the authenticity of the contents of advertisements.	Over the past few years, China has published several laws to regulate the use of AI. Of most noteworthy is the Measures for AI Generation and Synthetic Content Identification (2025), which requires both explicit identification to generated synthetic content (Article 4) and the addition of metadata (Article 5). This law applies to all AI generated synthetic content which includes text, pictures, audio, video, virtual scenes and other information generated and synthesised by AI technology (Article 3).
EU	European Consumer protection law centres around the Unfair Commercial Practices (UCPD 2005/29/EC) and the Misleading and Comparative Advertising (2006/114/EC) Directives and is the foundations of consumer protection law across individual Member States. Under the UCPD, the EU regards a commercial practice as being unfair if it materially distorts or is likely to distort the economic behaviour of the average consumer; and misleading if it contains false information, deceives or is likely to deceive the average consumer. The UCPD also contains a list of prohibited practices which include false endorsements and non-disclosure of paid promotions.	The EU's AI Act was the world's first, and with 144 pages, a comprehensive law that went into effect in 2024. The EU AI Act, which draws from EU product safety principles, only applies to AI systems placed on the market and categorises AI systems as high, medium and low risk. It also specifically defines the role of AI providers, deployers, importers and distributors. There is a small reference to advertising in Recital 29 which states: "In addition, common and legitimate commercial practices, for example in the field of advertising, that comply with the applicable law should not, in themselves, be regarded as constituting harmful manipulative AI-enabled practices."

[8] https://www.lexiscn.com/law/law-english-1-3999641-T.html?eng=0
[9] https://www.cac.gov.cn/2025-03/14/c_1743654684782215.htm

EU	Similarly, the Misleading and Comparative Advertising Directive prohibits misleading advertising which covers any advertising which in any way, including its presentation, deceives or is likely to deceive the persons to whom it is addressed or whom it reaches and which, by reason of its deceptive nature, is likely to affect their economic behaviour. It is worth noting that the EU's Digital Services Act refers to 'dark patterns', which are methods described as materially distorting or impairing consumer autonomous and informed decision-making, on online platforms in its recitals. It is expected that the upcoming Digital Fairness Act will specifically outlaw or restrict dark patterns. The vast majority of EU Member States have a consumer protection authority and an advertising self-regulatory body working in tandem to enforce and uphold consumer protection standards and advertising practices.	When drafting the law, EU policymakers saw the need for labelling or disclosure to meet transparency obligations of AI-generated or AI manipulated content (Recital 134). It also encourages the development of Codes of Practice to facilitate the detection and labelling of AI-generated or manipulated content (Recital 135). Furthermore, Article 50 sets out specific transparency obligations under Article 50 (2) and (4). Providers of AI systems that generate synthetic audio, image, video or text are required to mark the output in a way that is machine readable and detectable that it is artificially generated. Similarly, deployers of an AI system that generates or manipulates image, audio or video content constituting a deep fake, are required to disclose that the content has been artificially generated or manipulated. But critically, where the content forms part of an evidently artistic, creative, satirical, fictional or analogous work or programme, the transparency obligations are limited to disclosure of the existence of such generated or manipulated content in an appropriate manner that does not hamper the display or enjoyment of the work. Although Recital 29 suggests that advertising is excluded, the reference to deep fake in Article 50 (4) makes things slightly murky as it is defined in Article 3 (60) as being an AI-generated or manipulated image, audio or video content that resembles existing persons, objects, places, entities or events and would falsely appear to a person to be authentic or truthful. Strictly speaking this could also apply to the use of synthetic content in advertising; however, it is important not to conflate advertising, which has a specific commercial and creative purpose, with deep fakes. Deep fakes, a portmanteau of deep learning and fake, has a negative connotation and is frequently associated with malicious use.

India	India's Consumer Protection Act (CPA) 2019 establishes the Central Consumer Protection Authority (CCPA), which is the body that regulates misleading advertising and enforces consumer rights. It can impose penalties on manufacturers, service providers, and advertisers. The Advertising Standards Council of India is a self-regulatory body that promotes responsible advertising and maintains the advertising self-regulatory code to uphold standards. The CPA is the main law which regulates misleading advertising and unfair trading practices, and there are specific media and sector laws that regulate certain types of advertising. The CPA defines misleading advertising as an advertisement, which— • falsely describes such product or service. • gives false guarantees. • conveys an express or implied representation • deliberately conceals important information.	India currently does not have a dedicated AI regulatory framework and relies on existing laws to govern AI, such as the Digital Personal Data Protection Act 2023, Information Technology Rules 2021, and the Information Technology Act 2000. Information Technology Rules 2021 Part II Article 3 paragraph (2) b does refer to "artificially morphed images" in the context of deepfake pornography and mandating take down within 24 hours of receiving a complaint.
UK	Unfair commercial practices are prohibited by the Digital Markets, Competition and Consumer Act (DMCCA) which was passed in 2024 and came into force in 2025. A commercial practice is deemed unfair if it causes the average consumer to make a decision that they would not otherwise take if the practice involved a misleading action or misleading omission among others (Section 225). Both provisions are interpreted to apply to advertising. Section 225 defines a misleading action. Pertinent to this discussion is the consideration of the provision of false or misleading information relating to a product, a trader or any other matter relevant to a transactional decision; and, presenting something in manner that is likely to deceive the average consumer about a matter relating to a product, a trader or any other matter relevant to a transactional decision. In both cases, it would still count as misleading if it is presented in a misleading way even if the information is true.	The UK currently does not have a dedicated AI regulatory framework and relies on a sectoral approach and existing laws to govern AI. The UK published its white paper "A pro-innovation approach to AI regulation" under the Sunak government detailing its approach to AI governance. The UK did not make a specific pronouncement on its position on labelling, seemingly leaving that question for regulators to determine those requirements. In a paper called "Processes for Frontier AI Safety", the government outlined the need to distinguish synthetic content from content that was created by a human, and outlined three areas of practice to develop work in this area: • Research techniques that allow AI-generated content to be identified • Explore the use of watermarks for AI-generated content, including those that are robust to various perturbations • Explore the use of AI output databases

[10] https://consumeraffairs.nic.in/sites/default/files/CP%20Act%202019.pdf
[11] https://www.meity.gov.in/static/uploads/2024/02/Information-Technology-Intermediary-Guidelines-and-Digital-Media-Ethics-Code-Rules-2021-updated-06.04.2023-1-2.pdf
[12] https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach
[13] https://assets.publishing.service.gov.uk/media/653aabbd80884d000df71bdc/emerging-processes-frontier-ai-safety.pdf

India	Misleading omissions (Section 227) covers practices that • Omit material information • Omit information which the trader is required under any other enactment to give to a consumer as part of the practice, or • Fail to identify its commercial intent (unless it is already apparent from the context) The DMCCA also includes a list of 'banned practices', including creating a false impression the trader is not acting for purposes relating to the trader's business or falsely representing oneself as a consumer. This can sometimes be raised by regulators like the CMA in relation to the omission of labels such as 'Ad' from social media posts. The DMCCA grants the CMA new powers to directly enforce consumer legislation, alongside Trading Standards. The Advertising Standards Authority (ASA) is a self-regulatory body that promotes responsible advertising and maintains and enforces its advertising self-regulatory codes (CAP and BCAP).	
US	15 US Code Chapter 2 (Federal Trade Commission) §45prohibits unfair or deceptive acts or practices, including the dissemination of false advertisements; §52 defines false advertisements as ads for foods, drugs, devices, cosmetics, or services that are materially misleading. Section 5(a) grants the powers to the Federal Trade Commission to enforce the prohibitions on unfair or deceptive practices. 15 US Code Chapter 22 (Lanham Act) §1125 allows competitors and consumers to bring civil cases against misleading information that: • Is likely to cause confusion, or to cause mistake, or to deceive as to the affiliation, connection, or association of such person with another person, or as to the origin, sponsorship, or approval of his or her goods, services, or commercial activities by another person	Prior to it being rescinded, the President Biden era Executive Order (EO 14110) on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (AI) articulated the need to foster the development of effective labelling and content provenance mechanisms, so that citizens would be able to determine when content was generated using AI and when it is not. In addition, it called for US agencies to identify existing standards, tools, methods and practices with a view to improving them with a view to, for example:- • Authenticating content and tracking its provenance • Labelling synthetic content, such as using watermarking • Detecting synthetic content • Preventing child sexual abuse material or producing deepfake pornography • Auditing and maintaining synthetic content

[14] https://www.govinfo.gov/content/pkg/CFR-2024-title3-vol1/pdf/CFR-2024-title3-vol1-eo14110.pdf

	• In commercial advertising or promotion, misrepresents the nature, characteristics, qualities, or geographic origin of his or her or another person's goods, services, or commercial activities In addition to these Federal laws, individual States have additional rules that regulate advertising. BBB National Programs' National Advertising Division (NAD) and the National Advertising Review Board, established in 1971, is the US system of independent industry self-regulation to build consumer trust in advertising and support fair competition in the marketplace. NAD holds national advertising to high standards of truth and accuracy by reviewing truth-in-advertising challenges from businesses, trade associations, consumers, or on its own initiative.	H.R.3831 – AI Disclosure Act of 2023 introduced during the 118th Congress (2023-2024) was another initiative to introduce laws to mandate that all generative AI contain a disclosure along the lines of "Disclaimer: this output has been generated by artificial intelligence." This law did not receive a vote and fell away at the end of a previous session of Congress. While there is a myriad of State-level AI regulation, California has been one of the most prolific in signing AI laws into the statute book. It is arguably the most consequential State legislation given that all the leading AI companies are based in Silicon Valley. California's AI laws centre around the risks associated with Deep-fakes. For example, AB-2905 require robo-callers to disclose when deep fake generated voices are used. Laws such as AB-1831, SB-926, and SB-891 were drafted to protect individuals from exploitation via deepfake sexual imagery. Moreover, to protect electoral integrity, AB-2655, AB-2839, and AB-2355 intro-duce a several measures, such as platform labelling, removal and additional disclosure obligations placed on creators.

[15] https://www.congress.gov/bill/118th-congress/house-bill/3831/text

Singapore	Consumer protection is codified in Singapore's Consumer Protection (Fair Trading) Act 2003 (Rev 2020) (CPFTA). Consumers are protected from unfair practices which are defined as including the following actions: • Doing or saying anything, or omitting to do or say anything, which results in a consumer being deceived or misled • Making false claims; and • Taking advantage of consumers who are not in a position to protect their interests or is not reasonably able to understand the nature of the transaction Second Schedule Part 1 Specific Unfair Practices prohibits making a false or misleading representation concerning the need for any goods or services. The Competition and Consumer Commission of Singapore (CCCS) is the administering agency of the CPFTA, with investigative and enforcement powers. The Advertising Standards Authority of Singapore is the self-regulatory body that is responsible for maintaining the advertising code of practice. The code of practice, whilst non-binding, adds additional clarity to the CPFTA regarding advertising. According to section 5, advertisements should not mislead in any way by inaccuracy, ambiguity, exaggeration, omission or otherwise. In particular, it should not misrepresent any matter likely to influence consumers' attitude to any product, advertiser, or promoter.	Singapore does not have regulations specific to the governance of AI. Instead, it takes a sectoral approach to regulating aspects of AI. However, the government published its Model AI Governance Framework (Generative AI) in 2024. In the framework, it promotes the use of watermarking and cryptographic provenance to label and provide additional information to help flag content created or modified by AI.

[15] https://sso.agc.gov.sg/Act/CPFTA2003
[16] https://asas.org.sg/Portals/0/SCAP%202008_1.pdf

19. References

Altay,S. & Gilardi, F. (2024). People are sceptical of headlines labelled as AI-generated, even if true or human-made, because they assume full AI automation, PNAS Nexus Advance access publication, Volume 3, Issue 10, October 2024, pg403, https://doi.org/10.1093/pnasnexus/pgae403

Baek, T. H., Kim, J., & Kim, J. H. (2024). Effect of disclosing AI-generated content on prosocial advertising evaluation. International Journal of Advertising, 1–22. https://doi.org/10.1080/02650487.2024.2401319

Benway, J. P. (1998). Banner Blindness: The Irony of Attention Grabbing on the World Wide Web. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 42(5), 463-467. https://doi.org/10.1177/154193129804200504 (Original work published 1998).

Bradley, S. (2024). Top creatives share how they're grappling with AI use at agencies. The Drum [online]. Published 8 March 2024. Available from https://www.thedrum.com/news/2024/03/08/top-creatives-share-how-they-re-grappling-with-ai-use-agencies [Accessed 17 March 2024].

Campbell, C. L., Plangger, K., Sands, S., & Kietzmann, J. (2021). Preparing for an era of deepfakes and AI-Generated ads: A framework for understanding responses to manipulated advertising. Journal of Advertising Vol 51 No 1 Pages 22-38 (2022). Available from https://doi.org/10.1080/00913367.2021.1909515

Chakravorti, B. (2024). AI's Trust Problem. Harvard Business Review. Published 3 May 2024. Available from https://hbr.org/2024/05/ais-trust-problem [Accessed 14 March 2025].

Chesterman, S., 2025. Good models borrow, great models steal: intellectual property rights and generative AI. Policy and Society [Online], 44(1), pp.23–37. Available from: https://doi.org/10.1093/polsoc/puae006

Coca-Cola (2024). The Holiday Magic is Coming. Published 18 November. Available from https://www.youtube.com/watch?v=4RSTupbfGog [Accessed 10 March 2025].

Dempsey, M. (2021). Why artificial intelligence is being used to write adverts. BBC News [online]. Published 6 August 2021. Available from https://www.bbc.co.uk/news/business-57781557 [Accessed 17 March 2024].

Duffourc, M., Gerke, S. & Kollnig, K. (2024). Privacy of Personal Data in the Generative AI Data Lifecycle (July 08, 2024). Maastricht Faculty of Law Working Paper Forthcoming, Available from http://dx.doi.org/10.2139/ssrn.4899219

Echterhoff. G., Groll, S. & Hirst, W. (2007). Tainted truth: overcorrection for misinformation influence on eyewitness memory. Social Cognition. 25(3):367–409.

Edelman (2025). 2025 Edelman Trust Barometer Global Report – Trust and the Crisis of Grievance. Edelman Trust Institute. Available from https://www.edelman.com/sites/g/files/aatuss191/files/2025-01/2025%20Edelman%20Trust%20Barometer%20Global%20Report_01.23.25.pdf [Accessed 14 March 2025]

Epstein, Z., Fang, M.C., Arechar, A.A., & Rand, D.G. (2023). What label should be applied to content produced by generative AI? [online]. Available from: osf.io/preprints/psyarxiv/v4mfz_v1.

Feuerriegel, S., Hartmann, J., Janiesch, C. & Zschech, P. (2024). Generative ai. Business & Information Systems Engineering, 66(1), pp.111-126. Available from https://link.springer.com/content/pdf/10.1007/s12599-023-00834-7.pdf

Fisher, S.A. (2025). Something AI Should Tell You—The Case for Labelling Synthetic Content. Journal of Applied Philosophy, 42(1), pp.272-286. Available from https://onlinelibrary.wiley.com/doi/pdf/10.1111/japp.12758

Freeze, M., Baumgartner, M., Bruno, P., Gunderson, J.R., Olin, J., Ross, M.Q. & Szafran, J. (2021). Fake claims of fake news: political misinformation, warnings, and the tainted truth effect. Political Behaviour 43: 1433-1465. https://doi.org/10.1007/s11109-020-09597-3

Getty Images (2024). GettyImages VisualGPS – Building Trust in the Age of AI. Available from http://reports.gettyimages.com/VisualGPS-Building-Trust-In-AI.pdf. [Accessed 14 March 2025]

Jussupow, E., Benbasat, I., & Heinzl, A. (2020). Why are we averse towards algorithms? A comprehensive literature review on algorithm aversion. Research Papers. 168. https://aisel.aisnet.org/ecis2020_rp/168

Longoni, C., Fradkin, A., Cian, L., and Pennycook, G. (2022). News from Generative Artificial Intelligence Is Believed Less. In Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FACCT '22). Association for Computing Machinery, NY, USA, 97–106. https://doi.org/10.1145/3531146.3533077

McDermott, J. (2023). Balenciaga Pope: What an A.I.-generated puffer coat reveals about Francis' appeal (2023). https://www.americamagazine.org/arts-culture/2023/04/03/pope-francis-puffer-coat-245025 [Accessed 31 March 2025].

Mercer, S. & Watson, T. (2024). Generative AI in Cybersecurity: Assessing impact on current and future malicious software. CETaS Briefing Papers (June 2024). Available from https://cetas.turing.ac.uk/sites/default/files/2024-06/cetas_briefing_paper_-_generative_ai_in_cybersecurity.pdf [Accessed 27 May 2025].

Morrow, G., Swire-Thompson, B., Polny, J., Kopec, M., & Wihbey, J. (2020). The Emerging Science of Content Labeling: Contextualizing Social Media Content Moderation (December 3, 2020). Available at http://dx.doi.org/10.2139/ssrn.3742120.

Motamed, S., Culp, L., Swersky, K., Jaini, P. & Geirhos, R. (2025). Do generative video models learn physical principles from watching videos? arXiv preprint arXiv:2501.09038

Pennycook, G., Bear, A., Collins, E. T., & Rand, D. G. (2020). The implied truth effect: Attaching warnings to a subset of fake news headlines increases perceived accuracy of headlines without warnings. Management Science, 66(11), 4944–4957. https://doi.org/10.1287/mnsc.2019.3478

Roth, E. (2024). 'Google explains Gemini's 'embarrassing' AI pictures of diverse Nazis,' The Verge, 23 February. https://www.theverge.com/2024/2/23/24081309/google-gemini-embarrassing-ai-pictures-diverse-nazi [Accessed 31 March 2025].

Sweney, M. (2023). 'It's fundamental': WPP chief on how AI has revolutionised advertising. The Guardian [online]. Published 23 Feb 2023. Available from https://www.theguardian.com/technology/2023/feb/23/ai-artificial-intelligence-wpp-global-advertising-revolution-technology [Accessed 17 March 2024].

Tan, S. (2024a) AI-generated advertising: How do consumers around the world feel? YouGov. Published 5 June 2024. Available from https://business.yougov.com/content/49622-artificial-intelligence-ai-generated-brand-advertising-how-do-consumers-around-the-world-feel-comfortable. [Accessed 14 April 2025]

Tan, S. (2024b) Artificial Intelligence for Marketing: When should businesses disclose AI use for brand advertising? YouGov. Published 5 June 2024. Available from https://business.yougov.com/content/49623-artificial-intelligence-for-marketing-when-should-businesses-disclose-ai-use-for-brand-advertising. [Accessed 14 April 2025].

Thomas, D. (2024). How AI is transforming the business of advertising. FT.com [online]. Published 11 Feb 2024. Available from https://www.ft.com/content/0ac1fe79-4c2a-4030-8046-061293ba1127 [Accessed 17 March 2024]

Thomas, D. (2024a). WPP says 'too early' to know whether AI overhaul will hit advertising jobs. FT.com [online]. Published 30 Jan 2024. Available from https://www.ft.com/content/f649b685-bb09-4784-acfd-760ffbc53172 [Accessed 17 March 2024].

Thomas, D. (2024b). Publicis plans €300mn AI investment after exceeding growth targets. FT.com [online]. Published 25 Jan 2024. Available from https://www.ft.com/content/6cf203e3-2b5a-4740-8571-f57067f55f83 [Accessed 17 March 2024].

Toys R Us (2024). The Origin of Toys R Us: Brand Film Teaser | Toys R Us. Published 21 June 2024. Available at https://www.youtube.com/watch?v=F_WflzYGlg4 [Accessed 10 March 2025].

Williamson, D. A & Koch, J. (2024). The AI Ad Gap: Why Young Consumers Aren't Yet Buying Into Gen AI Ads. IAB. Published 9 December 2024. Available from https://www.iab.com/insights/why-young-consumers-avoid-gen-ai-ads/. [Accessed 14 April 2025].

Wittenberg, C., Epstein, Z., Berinsky, A. J., & Rand, D. G. (2023). Labeling AI-Generated Content: Promises, Perils, and Future Directions. MIT Schwarzman College of Computing. Available from https://computing.mit.edu/wp-content/uploads/2023/11/AI-Policy_Labeling.pdf [Accessed 16 March 2025]

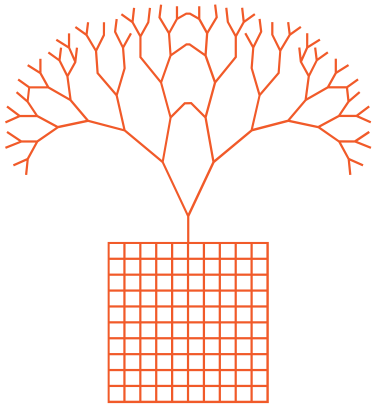
WPP (2022). Ogilvy and Wavemaker: #NotJustACadburyAd and Shah Rukh Khan-My-Ad. Wpp.com [online]. Available from https://www.wpp.com/featured/work/2022/03/ogilvy-and-wavemaker-notjustacadburyad [Accessed 18 March 2024].

Yao, J.Y., Ning, K.P., Liu, Z.H., Ning, M.N., Liu, Y.Y. & Yuan, L. (2024). LLM lies: Hallucinations are not bugs, but features as adversarial examples. arXiv preprint arXiv:2310.01469.

Zhou, J., Zhang, Y., Luo, Q., Parker, A.G. and De Choudhury, M., 2023. Synthetic Lies: Understanding AI-Generated Misinformation and Evaluating Algorithmic and Human Solutions. Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems [Online], CHI '23: CHI Conference on Human Factors in Computing Systems, Hamburg Germany. Hamburg Germany: ACM, pp.1–20. Available from: https://doi.org/10.1145/3544548.3581318 [Accessed 27 May 2025].

For more information about the ICAS Global Think Tank, please contact Sibylle Stanciu, ICAS Director (sibylle.stanciu@icas.global)

The International Council for Advertising Self-Regulation (ICAS)
https://icas.global



ICAS
GLOBAL
THINK
TANK