

AI Hallucinations: When Creation Comes at a Cost, Who Pays?

[DSK Legal](#)

[Australia, China, European Union, India, Singapore, United Kingdom, USA](#)

April 29 2025

Artificial Intelligence has transformed the way information is processed and consumed, offering unprecedented capabilities and value across fields like medicine, law, journalism, and finance. However, alongside its advancements, AI models have been riddled with technological and regulatory challenges. One such conundrum has been regulation of AI hallucinations. **AI hallucination occurs when Generative AI models (GenAI) produce factually inaccurate or misleading responses, and stem broadly from biases in or quality of training data and model limitations.**

AI models may be categorised as intermediaries in instances like AI-powered search engines that retrieve and summarize third-party content, chatbots that transmit pre-existing information without independent analysis, and content recommendation systems that rank user-generated content without modification. Such platforms may be protected under safe harbour provisions given their role as a mere conduit. **However, GenAI operates differently – it independently generates responses by self-learning from existing datasets, blurring liability lines. From fabricating legal cases to generating false news, GenAI poses risks of misinformation.** In certain cases, AI hallucinations may pose a cybersecurity threat by misleading developers into deploying codes which may compromise systems. In 2023, Alphabet Inc lost \$100 billion in market value after Bard shared inaccurate information in a promotional video exemplifying the potential fallout of GenAI. In sensitive sectors such as healthcare, law, and finance, such hallucinations can mislead diagnoses, distort legal arguments, and impact financial decisions, leading to reputational damage, regulatory scrutiny, and legal liability. Another recent illustration is the order issued by the Bengaluru bench of the Income Tax Appellate Tribunal (ITAT) in *Buckeye Trust v. Principal Commissioner of Income Tax* (ITA No. 1051/Bang/2024) which relied on four fictitious legal precedents generated by an AI tool.

The Liability Web

Determining liability for AI hallucinations is a challenge – since GenAI models rely on probabilistic reasoning, their responses are neither fully deterministic nor easily traceable. The black-box nature of large language models (LLMs) further hampers determination of accountability. Ambiguous or low-quality data is another key contributor to such hallucinations. As AI becomes integral to decision-making, a crucial question arises – who is responsible when AI-generated misinformation leads to real-world consequences? Should accountability rest with developers, the organizations deploying these systems, or users relying on their outputs? If misinformation stems from flawed training datasets rather than a design flaw, should liability rest with developers, data sources, or platform providers?

At the heart of this issue is the debate over responsibility. AI companies argue that hallucinations are an inherent limitation of GenAI models, placing the burden on users to verify the responses/ outputs. While regulators are exploring frameworks to hold developers accountable for accuracy in their system generated outputs, reactive fixes may be insufficient and AI regulation will need to focus on striking a balance between the mechanics of the model and the impact of the output. For instance, the EU AI Act classifies AI systems by risk level, enforcing stringent accuracy requirements for high-risk AI applications in legal and healthcare systems. This approach was evident in the shutdown of *Tessa*, an AI-powered mental health chatbot that provided harmful advice to users with eating disorders. However, this approach has not been consistent globally. The US AI Bill of Rights adopts a more user-centric approach, emphasizing on verification by users rather than developer liability, and limiting developer liability only to the extent of transparency, content labelling, and safety measures for advanced AI systems – a stance tested when two lawyers and a law firm were fined for citing fake cases generated by ChatGPT in a court filing. Other countries, such as the UK, China, Australia, and Singapore, are refining oversight through sectoral regulations, non-binding guidelines, and targeted compliance measures, focusing on explainability, accountability, and safeguards against AI hallucinations and AI-generated misinformation.

Responsible AI Deployment – A Shared Responsibility

Given the wide range of complexities, it seems that a shared-responsibility regulatory model is emerging, distributing liability for AI hallucinations between developers, deployers and users based on factors such as control, foreseeability, and negligence. From a developer's perspective, AI regulations should prioritize the adoption of best practices to mitigate hallucinations, without becoming overly prescriptive. These best practices should broadly centre on rigorous dataset validation—ensuring that models are trained on high-quality, diverse, and bias-mitigated data—and the implementation of regular audits to ensure models remain responsive and adaptive to evolving risks. From a deployer standpoint, key mitigation measures may include mandatory disclosure and warning labels, informing users about AI limitations and error rates and implementing human-in-loop verifications. Additionally, users may be provided with an option to flag hallucinated responses, triggering a review by a fact-checking unit for dataset corrections. From a user's perspective, it is essential to exercise due diligence and avoid negligent reliance on AI systems. Being informed, cautious, and deliberate in how AI outputs are interpreted and applied can help prevent unintended consequences—especially in high-stakes or professional contexts. Together, enforcing layered responsibilities reflects a more nuanced and resilient approach towards governing AI hallucinations - one that acknowledges the dynamic interaction between technology and its human handlers. By aligning incentives and accountability across the AI lifecycle, a shared-responsibility model fosters both innovation and trust in the safe, ethical deployment of GenAI.