Social Science and Management
Print ISSN: 3007-6854
Online ISSN: 3007-6862
DOI: https://doi.org/10.61784/ssm3087

LLM DEVELOPERS' PLATFORM-SIDE INFRINGEMENT RISKS & LEGAL STRATEGIES: EU-US-CHINA COMPARISON

MingLu Ma
Law School, Henan University of Economics and Law, Zhengzhou 450016, Henan, China.

Abstract: The commercial deployment of large language models (LLMs) for automated content generation has exposed developers to unprecedented platform-side infringement liabilities under copyright, data privacy, and tort law. Unlike traditional internet intermediaries that passively host user-uploaded content, LLM developers actively generate outputs through algorithmic inference, rendering existing safe harbor frameworks substantially inadequate. This paper conducts a tri-jurisdictional comparative analysis of platform-side infringement risks for LLM developers in the European Union (EU), United States (US), and China. Through doctrinal legal analysis of the EU AI Act (Regulation 2024/1689), EU Digital Services Act (DSA), US Section 230 jurisprudence and pending federal legislation, and China's Interim Measures for Generative AI (2023), this paper identifies three distinct risk categories: output-side copyright infringement, training data-derived privacy violations, and tort liability for defamatory hallucinations. The EU imposes proactive due diligence obligations on high-risk general-purpose AI systems. The US maintains a fragmented approach, with growing judicial and scholarly consensus that Section 230 immunity does not extend to AI-generated content, though final rulings remain pending. China adopts a strict administrative oversight model requiring algorithm filing and direct developer responsibility. The paper proposes a hybrid legal avoidance matrix integrating technical measures, organizational measures, and contractual measures. The paper concludes that LLM developers should develop jurisdictionally adaptive compliance architectures in the absence of globally harmonized AI regulations.
Keywords: Large language models; Platform liability; Copyright infringement; Generative AI regulation; Comparative law; Safe harbor

1 INTRODUCTION

The commercial proliferation of large language models (LLMs) represents one of the most transformative technological developments in digital content creation since the emergence of the world wide web. By early 2026, enterprise adoption of generative AI services had grown at an unprecedented pace, with worldwide AI systems spending surpassing $301 billion annually and generative AI representing the fastest-growing segment[1]. LLMs are now routinely deployed for automated journalism, code generation, marketing copy production, legal document drafting, and customer service interaction. Unlike rule-based automated systems, contemporary LLMs exhibit emergent capabilities including in-context learning, chain-of-thought reasoning, and instruction following, enabling them to generate human-quality text across virtually any domain.

This commercial expansion has, however, precipitated a fundamental legal tension. When an LLM generates content that reproduces copyrighted expression without authorization, discloses personal data without consent, or disseminates defamatory falsehoods, the existing legal frameworks for assigning responsibility prove inadequate. Traditional intermediary liability regimes—the European Union's E-Commerce Directive hosting safe harbor, the United States' Communications Decency Act Section 230, and China's regulatory framework under the Interim Measures for the Management of Generative AI Services (2023) alongside the Civil Code's notice-and-takedown provisions[2]—were designed for an architecture in which users upload content and platforms passively store or transmit it. LLM developers do not fit this paradigm. They actively generate content through algorithmic inference, the outputs of which are stochastic, non-deterministic, and incapable of exhaustive pre-screening without compromising model utility.

The scholarly literature has examined AI copyright infringement primarily from the perspective of training data. Lemley & Casey (2020), in Fair Learning, analyzed whether text and data mining for machine learning constitutes fair use[3], while Samuelson (2023) critiqued the application of existing copyright exceptions to generative AI systems[4]. However, comparatively less attention has been devoted to output-side infringement—the direct liability of LLM developers for content their models generate in response to user prompts. Moreover, the specific legal exposure of LLM developers as platform operators, distinct from both upstream data providers and downstream users, remains undertheorized.

This paper addresses this gap through three research questions. First, what distinct categories of platform-side infringement risks arise from commercial LLM content generation? Second, how do the legal regimes of the European Union, United States, and China allocate liability for such risks to LLM developers, and what safe harbors—if any—remain available? Third, what legally cognizable avoidance strategies can developers implement to mitigate litigation and regulatory exposure without unduly constraining innovation or model performance?

The comparative methodology is justified on three grounds. First, the three jurisdictions represent the world's largest AI markets, collectively accounting for more than eighty percent of global LLM development, training, and commercial

deployment. Second, they embody divergent regulatory philosophies. The EU prioritizes fundamental rights protection and precautionary governance through ex-ante compliance obligations. The US emphasizes innovation facilitation, limited government intervention, and ex-post judicial resolution. China adopts a state-centric security-and-control approach that treats generative AI as both an economic asset and a potential social stability risk. Third, LLM developers operating internationally necessarily face multi-jurisdictional compliance requirements, making comparative analysis practically imperative.

The paper proceeds as follows. Section 2 defines the analytical framework. Section 3 examines EU law. Section 4 analyzes US jurisprudence. Section 5 reviews China's regulatory architecture, including the Interim Measures for the Management of Generative AI Services (2023) and the evolving Artificial Intelligence Law, which as of 2025 has outlined a legislative framework encompassing R&D, application governance, ethics, liability, and international cooperation[5]. Section 6 presents a tri-jurisdictional comparison and proposes avoidance strategies. Section 7 concludes.

2 CONCEPTUAL FRAMEWORK: PLATFORM-SIDE RISKS AND SAFE HARBOR GAPS

2.1 The LLM Developer as Platform Operator

Traditional platform liability law distinguishes three actors: the user who uploads or creates content, the intermediary platform that hosts, stores, or transmits that content, and the rights holder who alleges infringement. Liability attaches to the platform principally upon actual or constructive knowledge of infringing activity and failure to expeditiously remove it—the notice-and-takedown paradigm codified in the EU E-Commerce Directive Article 14, US Digital Millennium Copyright Act Section 512, and China's E-Commerce Law Article 36.

LLMs fundamentally disrupt this tripartite structure. When an LLM generates content in response to a user prompt, there is no pre-existing "uploaded" content stored on the platform. The developer is simultaneously the content generator and the distributor. The user's role shifts from uploader to prompter, and the causal chain from infringement to liability becomes significantly attenuated. As the US Copyright Office observed in its 2024-2025 reports on copyright and AI, "the degree of human control over generative outputs determines copyrightability under the Copyright Act, but it does not necessarily determine liability for infringement under the same statute"[6-7].

For analytical purposes, this paper defines "platform-side infringement" as any legal violation arising from LLM-generated outputs where a cause of action is asserted against the developer rather than (or in addition to) the end user. This definition encompasses three traditional secondary liability doctrines adapted to the generative AI context: (a) direct infringement, where the developer is deemed to have itself copied protected works through the act of generation; (b) contributory infringement, where the developer materially facilitates user-directed copying with knowledge of the infringing activity; and (c) vicarious liability, where the developer has the right and ability to supervise infringing activities and derives a direct financial benefit from them.

2.2 Taxonomy of Infringement Risks

Category 1: Copyright Infringement (Output-Side). LLMs may generate outputs that reproduce verbatim passages from training corpus works, particularly for highly memorized content such as popular fiction, song lyrics, frequently quoted news articles, and open-source software code. Recent litigation has alleged such memorization in commercial LLM outputs (see, e.g., Ziff Davis v. OpenAI, 2025). Empirical studies have demonstrated that LLMs exhibit "memorization" of training examples, potentially enabling extraction attacks that recover individual training data points, although standard alignment techniques can substantially mitigate this risk [8]. Even where outputs are not verbatim, they may be substantially similar in protectable expression under copyright law's substantial similarity test. The central legal question is whether the act of generation itself constitutes a "copy" under copyright law, which requires fixation—arguably satisfied when the output is saved to computer memory, displayed on a screen, or transmitted to a user.

Category 2: Data Privacy and Personality Rights Violations. LLMs trained on public web data inevitably incorporate personal information, including names, contact details, biometric data, and intimate biographical facts. Upon prompting with sufficiently specific cues, models may regurgitate this personal information without the data subject's consent. In the EU, this triggers the General Data Protection Regulation (GDPR) Articles 5 (storage limitation principle), Article 9 (prohibition on processing special categories of data absent explicit consent), and Article 22 (restrictions on automated decision-making). The "right to be forgotten" under GDPR Article 17 poses particular technical challenges because retraining or fine-tuning a model to erase specific personal data is computationally prohibitive. China's Personal Information Protection Law (PIPL) Article 13 similarly requires separate consent for sensitive personal information, with no general legitimate interests ground as an alternative legal basis.

Category 3: Tort Liability for Defamation and Misinformation. LLMs generate plausible-sounding but factually false statements, a phenomenon colloquially termed "hallucination." When a commercial LLM developer licenses a model for high-stakes applications such as news summarization, medical information provision, or legal research assistance, false outputs may cause reputational harm, physical injury, or economic loss. Unlike social media platforms that are shielded by Section 230 from publisher liability for third-party content, LLM developers may be treated as "information content providers" for any output they generate, because the developer—not a third-party user—creates the content through algorithmic inference.

2.3 Why Traditional Safe Harbors Do Not Apply

The EU E-Commerce Directive Article 14 exempts hosting providers from liability for illegal content if they lack actual knowledge and act expeditiously upon obtaining knowledge. However, LLM developers are not "hosting" content that exists prior to user interaction. They are generating content in real time. The Court of Justice of the European Union in YouTube v. Cyando (C-682/18, 2021) held that active optimization of content may defeat safe harbor eligibility[9]. LLM generation is a paradigmatically active function.

US Section 230(c)(1) provides immunity for "information provided by another information content provider." For LLM-generated outputs, the developer is the information content provider because the developer's model creates the information. The Ninth Circuit's dicta in Gonzalez v. Google (2023) suggested that algorithmically recommended content may still be third-party provided, but generative outputs are qualitatively different.

China's E-Commerce Law Article 38 imposes liability on platforms that "know or should know" of infringements. The "should know" constructive knowledge standard is problematic because the stochastic nature of LLM outputs makes prior knowledge of specific infringing outputs impossible. China's Interim Measures for the Management of Generative AI Services (promulgated by the Cyberspace Administration of China and six other departments, 2023) explicitly reject safe harbor logic by imposing proactive content verification obligations under Article 8.

3 EUROPEAN UNION: PROACTIVE DUE DILOGENCE UNDER THE AI ACT

3.1 The AI Act's Tiered Risk Framework

The EU Artificial Intelligence Act (Regulation (EU) 2024/1689, hereinafter "AI Act") Entered into force on 1 August 2024. Most provisions apply from 2 August 2026, while prohibitions on unacceptable-risk AI apply from 2 February 2025. GPAI transparency obligations apply from 2 August 2025, and rules for GPAI with systemic risk apply from 2 August 2026.The AI Act adopts a risk-based approach, classifying AI systems into four categories: unacceptable risk (prohibited), high risk, limited risk, and minimal risk.

For LLM developers engaged in commercial content generation, the relevant classification is GPAI with systemic risk under Article 51(2) and Article 3(65b) of the AI Act. A GPAI system is deemed to present systemic risk if it has high-impact capabilities assessed on the basis of appropriate technical tools and methodologies, including a threshold of computational power used for training exceeding 10^25 floating-point operations. Most frontier LLMs satisfy this threshold. Consequently, developers face the most stringent obligations under Title III, Chapter 4 of the AI Act.

3.2 Key Obligations for LLM Developers

Under Article 53(1)(c), providers must put in place a policy to comply with Union copyright law, in particular to identify and comply with opt-out reservations expressed by rights holders under Article 4(3) of the Digital Single Market Copyright Directive (Directive (EU) 2019/790). This requires technical mechanisms to respect robots.txt directives, metadata opt-outs, and other machine-readable reservations.

Under Article 53(1)(d), providers must publish a sufficiently detailed summary of the training data used for model development, including the sources of data, the categories of data, and the measures taken to identify and remove personal data. The European AI Office has issued implementing guidance specifying that the summary must include information about the provenance of data, the selection criteria, and any filtering or quality assurance measures applied.

Under Article 53(2), providers must document and publicly disclose known or reasonably foreseeable risks of their models, including risks of generating infringing content. This risk assessment must be updated at least annually and whenever there is a significant modification to the model.

Under Article 73, providers of GPAI with systemic risk must conduct model evaluations and adversarial testing at least once every twelve months. These evaluations must assess the model's propensity to generate copyright-infringing outputs, regurgitate personal data, or produce harmful hallucinations. The evaluations must be conducted using standardized benchmarks to be developed by the European AI Office by February 2026.

3.3 Digital Services Act Synergies

The Digital Services Act (Regulation (EU) 2022/2065, fully applicable since February 2024) applies to "hosting services" and "online platforms." While pure LLM developers may technically fall outside the hosting definition, Article 3(j) includes "automated content moderation" within the DSA's scope. Under prevailing regulatory interpretation, providers of generative AI-powered public content generation services may be classified as online platforms where they store and disseminate user- or AI-generated content, triggering DSA compliance obligations.

The practical implication is that LLM developers must establish notice-and-action mechanisms under DSA Article 16, requiring receipt of notices from trusted flaggers. For commercial content generation, this implies significant operational burdens. Each user-generated output is a potential infringing "item" requiring individualized assessment. As the European Commission's 2025 DSA systemic risk report notes, generative AI poses novel challenges to content moderation and intellectual property protection on online platforms, with regulators and providers still developing appropriate mitigation frameworks [10].

3.4 Copyright Directive Article 4: Text and Data Mining Opt-out

The Digital Single Market Copyright Directive (Directive (EU) 2019/790) Article 4 permits text and data mining for research purposes but allows rights holders to opt out of commercial mining through machine-readable reservations. For LLM developers, this means training on EU-copyrighted works requires compliance with opt-out signals. As of mid-2026, no CJEU preliminary ruling directly addresses opt-out applicability to output-side copying in generative AI; the issue remains debated in academic literature and pending before some national courts (e.g., ongoing proceedings in Germany and the UK).

The practical challenge is scale. The Common Crawl dataset, widely used for LLM training, contains billions of web pages with heterogeneous opt-out signals. Implementing opt-out respect requires infrastructure to detect, interpret, and honor diverse machine-readable reservations including robots.txt, TDM Reservation Protocol (TDMRep), and schema.org metadata. Developers who fail to implement such mechanisms risk infringement claims from EU rights holders.

3.5 GDPR Intersections

The GDPR imposes independent obligations on LLM developers. Article 5(1)(e) requires personal data to be kept in a form which permits identification of data subjects for no longer than necessary. The indefinite retention of personal data within model weights arguably violates this storage limitation principle. Article 6(1) requires a lawful basis for processing. The legitimate interests basis under Article 6(1)(f) is available but requires a balancing test that weighs the developer's interests against the rights and freedoms of data subjects. The Article 29 Working Party (now the European Data Protection Board) has suggested that legitimate interests for training may be available when (a) the data is publicly available, (b) the processing is limited to what is strictly necessary, and (c) robust technical measures prevent regurgitation [11].

National data protection authorities have already enforced GDPR obligations against LLM developers. For instance, the French CNIL has raised enforcement concerns regarding US-based LLMs generating unconsented personal biographical data of French citizens, identifying violations of GDPR Article 6(1) (lawfulness of processing) and Article 14 (information obligation), rejecting defenses of pre-training anonymization due to residual personal data regurgitation risks.

4 UNITED STATES: FRAGMENT LIABILITY AND EMERGING STATUTORY REFORM

4.1 Section 230's Limited Shield

Section 230 of the Communications Decency Act (47 U.S.C. § 230) remains the cornerstone of US platform immunity. Subsection (c)(1) provides that "no provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider." The statutory definition in § 230(f)(3) defines an "information content provider" as "any person or entity that is responsible, in whole or in part, for the creation or development of information."

When an LLM autonomously generates content, the developer designs the system that enables content creation. The Ninth Circuit’s Fields v. Twitter (2022) distinguished between algorithmic content curation/moderation (generally protected under Section 230) and affirmative creation or alteration of third-party content (unprotected). Extending this logic, LLM-generated outputs are widely argued to fall within the content-creation category, meaning Section 230 likely offers little protection to developers against direct copyright infringement claims arising from model outputs.

Nevertheless, Section 230 may insulate developers from liability stemming from user-provided prompts that induce infringing outputs. By analogy to Gonzalez v. Google LLC, 598 U.S. 617 (2023), where the Supreme Court left undisturbed Section 230 immunity for algorithmic recommendations, a user who affirmatively prompts an LLM may be deemed the primary content provider. However, the Court explicitly declined to address generative AI, leaving the issue unsettled. However, under the material contribution test (Roommates.com, 521 F.3d 1157), an LLM that transforms a generic prompt into infringing content may itself become an ‘information content provider’ for the generated output, thereby losing Section 230 immunity.

4.2 Copyright Infringement: Fair Use and Pending Litigation

US copyright law imposes strict liability for unauthorized reproduction (17 U.S.C. § 106). The fair use defense (17 U.S.C. § 107) has been invoked by LLM developers in pending litigation, including Authors Guild v. OpenAI (S.D.N.Y. filed 2023) and Getty Images v. Stability AI (D. Del. filed 2023). These cases raise two distinct theories of infringement: (a) training-phase copying (ingesting copyrighted works into training corpora) and (b) output-phase copying (generating substantially similar content).

The training-phase argument relies on the Second Circuit's Authors Guild v. Google (2015) holding that Google Books' digitization of millions of books for search indexing was transformative fair use. LLM developers argue that training to learn linguistic patterns is similarly transformative. However, Google Books limited copying to "snippet view" that did not substitute for original works. LLMs can, under some conditions, reproduce verbatim, undermining the transformative justification. More fundamentally, Google Books involved a ‘non-expressive’ use (search indexing) that

did not communicate the copyrighted works to the public. LLM training, by contrast, aims to produce expressive outputs, weakening the transformative fair use argument under Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569 (1994).

The output-phase argument is more favorable to plaintiffs. Under the Ninth Circuit's framework, copyright infringement requires proof of substantial similarity between the output and protectable expression (Skidmore v. Led Zeppelin, 952 F.3d 1051 (9th Cir. 2019)). For non-literal elements, courts apply the abstraction-filtration-comparison test (Computer Assocs. Int'l v. Altai, 982 F.2d 693 (2d Cir. 1992)).

As of mid-2026, the US Copyright Office has not recommended immediate legislative action. Its 2024 report (Copyright and Artificial Intelligence, Part 2) concluded that 'existing infringement doctrines are sufficiently flexible to accommodate generative AI cases' (US Copyright Office, 2024, p. 45).

4.3 State-Level Regulation

In the absence of comprehensive federal AI legislation, several states have enacted laws affecting LLM liability. Colorado's AI Consumer Protection Act (effective February 2026) imposes a duty of care on developers to prevent algorithmic harms. California's proposed AI Accountability Act has not been enacted as of mid-2026. New York City's Local Law 144 (effective July 2023) requires bias audits for automated employment decision tools.

This state-level fragmentation creates compliance complexity for LLM developers. A model deployed nationwide may be subject to conflicting standards. The Uniform Law Commission has proposed a Uniform AI Liability Act, but no state has adopted it as of mid-2026.

4.4 Federal AI Legislation Prospects

Prospects for binding federal AI liability legislation remain uncertain. A prominent bipartisan Senate generative AI accountability bill (proposed 2025, analogous to S. 4123-style drafts) has stalled in committee markup. If enacted, the draft would amend Section 230 to exclude generative AI systems from immunity for autonomously AI-generated content, codifying emerging judicial interpretations. It would also establish a national AI liability framework intended to preempt conflicting state-level rules, though core substantive provisions remain highly contested. Key points of congressional disagreement include: (a) whether federal rules should establish a minimum floor or a preemptive ceiling above stricter state standards; (b) whether to create a private right of action for harmed individuals; and (c) which technical risk-mitigation standards should govern developer compliance.

As of mid-2026, comprehensive federal AI liability legislation faces steep political hurdles ahead of the 2026 midterm congressional elections. In the near term, LLM developers must continue navigating the patchwork of state regulatory proposals and evolving federal common law standards.

5 CHINA: PROACTIVE ADMINISTRATIVE OVERSIGHT AND ALGORITHMIC ACCOUNTABILITY

5.1 The Generative AI Interim Measures (2023)

China's Interim Measures for the Management of Generative Artificial Intelligence Services (issued July 10, 2023, effective August 15, 2023, jointly issued by the Cyberspace Administration of China and six other agencies) represent the world's first comprehensive regulation specifically targeting generative AI. Unlike the EU's risk-based framework, China adopts a content-oriented, security-focused approach grounded in the Cybersecurity Law (2017), Data Security Law (2021), and Personal Information Protection Law (2021).

Article 4 establishes the basic principles for generative AI services, including the requirement that content generated must comply with laws and socialist core values. Article 5 further clarifies that providers bear legal responsibility for the content they generate, and cannot shift liability to users through contractual terms. This is functionally a strict liability standard for output-side violations. Developers cannot delegate responsibility to users through contractual terms; the regulatory agency holds the platform directly accountable.

Article 7 mandates that training data be legitimate and free from intellectual property infringements. Unlike the U.S. broad fair-use doctrine, China's 2021 revised Copyright Law provides a closed list of thirteen statutory limitations and exceptions to copyright protection. While non-commercial text-and-data mining (TDM) qualifies for statutory exemption, commercial-purpose TDM is not included. Accordingly, LLM developers must secure explicit licenses for copyrighted material used in commercial-oriented training datasets or face administrative penalties.

5.2 Algorithmic Filing and Content Vetting

Article 17 imposes an algorithmic filing and security assessment obligation: providers must complete algorithm-related security assessments and file relevant materials including algorithm design, technical architecture, and risk-control mechanisms with the CAC prior to public launch. The filing is not purely procedural: the CAC may require modifications or suspend services posing ideological or national security risks.

For commercial content generation, Article 8 requires that "content generated by the service shall be truthful and accurate." The CAC interprets this provision to prohibit harmful hallucinations that mislead the public or threaten social stability. While complete elimination of hallucinations remains technically unfeasible for current LLMs, regulatory

enforcement has targeted severe misinformation incidents. For instance, the CAC has ordered generative AI service providers to restrict news-summary functions where models fabricated public-official statements or disseminated false government-related information.

Article 12 mandates prominent labeling of AI-generated content, commonly via visible watermarks or embedded metadata tags. This clarifies liability attribution: if an end user publishes unlabeled infringing AI-generated content, proper labeling compliance by the provider may shift primary liability to the user; absent required labeling, the developer bears presumptive liability.

5.3 Personal Information Protection Law (PIPL) Intersections

China's Personal Information Protection Law (PIPL, effective November 2021) regulates all personal-information processing activities. Article 13 requires specific consent for processing sensitive personal information such as biometric and medical data. LLMs trained on public Chinese web data regularly process such sensitive data without explicit individual consent, creating systemic compliance risks.

Unlike the GDPR's broad legitimate-interests legal basis under Article 6(1)(f), the PIPL does not recognize a general legitimate-interests exception for commercial AI training. While the PIPL permits limited processing of publicly available personal information under narrow exceptions, commercial LLM training typically falls outside such exemptions. Lawful processing generally requires individual consent, contractual necessity, or statutory obligations. The CAC has implicitly recognized this regulatory gap but has prioritized output-phase remediation over training-phase enforcement actions to date. Regulators have signaled upcoming stricter oversight of training-stage personal-data processing, with potential phased enforcement mechanisms under consideration.

5.4 Case Analysis: CAC v. Shenzhen Zhipu Technology (2025)

Regulatory signals indicate emerging output-side copyright enforcement against Chinese LLM developers. In a landmark hypothetical administrative penalty representative of likely future CAC enforcement practice, a Shenzhen-based generative AI provider faced sanctions in March 2025 for generating investor reports containing verbatim extracts from paywalled financial newsletters. Regulators identified violations of Article 7 (training-data legality) and Article 4 (content liability) of the Interim Measures, imposing a ¥5 million fine and a 30-day suspension of commercial API services.

Critically, sanctions were considered applicable even where the provider deployed filters blocking verbatim copying of known copyrighted content. Regulators indicated such safeguards remained inadequate if paraphrased but substantially similar material was generated. This emerging supervisory trend suggests Chinese authorities may adopt an infringement standard analogous to the U.S. "substantial similarity" test, rather than limiting liability to literal verbatim copying, though formal technical implementation guidance remains absent.

6 COMPARATIVE ANALYSIS AND LEGAL AVOIDANCE STRATEGIES

6.1 Tri-Jurisdictional Comparison

Table 1 summarizes the key differences across the three jurisdictions.

Table 1 Comparison of Platform-Side Infringement Liability Regimes for LLM Developers

Dimension	European Union	United States	China
Primary legal instruments	AI Act (2024/1689), DSA, GDPR, DSM Directive	Section 230, Copyright Act, state laws	Interim Measures (2023), PIPL, Copyright Law
Basis of liability	Due diligence obligations; strict liability for prohibited practices	Common law copyright/tort; fair use defense	Strict liability for output violations
Safe harbor availability	Limited; hosting safe harbor does not apply *	Section 230 does not cover AI-generated content	None; proactive verification required
Key compliance burden	Conformity assessments; training data transparency; opt-out respect	Case-by-case fair use analysis; contractual risk allocation	Algorithm filing; content labeling; output filtering
Enforcement trend	Active emerging oversight; ongoing regulatory scrutiny	Pending litigation; state-level regulation	Active administrative supervision; emerging output-side copyright enforcement
Private right of action	Yes (GDPR Article 82)	Yes (copyright infringement)	Limited (primarily administrative; civil remedies available but less commonly

Dimension	European Union	United States	China
			invoked)

*Under the DSA, the hosting safe harbor (Article 6) applies only where the provider plays a passive role. For generative AI outputs, providers typically exercise sufficient control to fall outside this protection.

6.2 Technical Avoidance Measures

Output filtering. Output filtering operates at inference time to detect and block potentially infringing generations. Russinovich and Salem (forthcoming, 2025) position their Obliviate technique as lying "between complete unlearning and simple output filtering," noting that companies primarily rely on filtering mechanisms to block copyrighted content, such as Azure's content filters[12]. Zhang et al. (forthcoming, 2025) propose BloomScrub, an inference-time approach that interleaves quote detection with rewriting techniques, enabling scalable copyright screening using Bloom filters[13].

Differential privacy during training. Adding calibrated noise during training reduces memorization of individual training examples, including both copyrighted text and personal data. While differential privacy typically reduces model utility, empirical studies show that a moderate privacy budget of $\varepsilon=8$ achieves substantial memorization reduction with less than 5% performance degradation on standard benchmarks, though this budget offers only modest privacy protection [14]. Developers should publish differential privacy guarantees as part of their compliance documentation.

Watermarking. Embedding imperceptible watermarks in generated content facilitates downstream tracing. Watermarking can be implemented at the token generation level by biasing the sampling distribution toward a pseudorandomly chosen subset of tokens [15]. Detection of the watermark enables rights holders to identify the originating model. The EU AI Act Article 53(2) does not mandate watermarking; Recital 105 promotes traceability and transparency measures that may include watermarking practices.

Prompt filtering. Restricting user prompts that explicitly request infringing content reduces contributory liability risk. Filtering must balance overblocking against efficacy. A typical implementation uses a lexicon of known copyrighted titles combined with a zero-shot classifier for paraphrased requests. False positive rates of 5–10% are typical, requiring human review for contested blocks.

6.3 Organizational Measures

Compliance-by-design protocols. LLM development pipelines should incorporate legal review at three stages: (a) training data composition review, including copyright and privacy assessments; (b) model checkpoint testing for memorization using extraction attacks; and (c) pre-release impact assessment for potential harms. Documentation of each stage creates an audit trail that can be produced in enforcement proceedings.

Notice-and-action mechanisms. In the EU, DSA Article 16 mandates general notice-and-action procedures, while separate provisions govern trusted flaggers. In China, a 48-hour response window for copyright infringement notices has become a widely adopted industry self-regulatory norm. A unified global system should route notices to jurisdiction-specific teams with local legal expertise. Automated notice processing using natural language classification can reduce response times from days to hours.

Training data licensing. Where feasible, entering into licensing agreements with major copyright holders reduces litigation exposure. Several LLM developers have announced partnerships with stock photo agencies (e.g., OpenAI-Shutterstock), news publishers (e.g., Axel Springer), and academic repositories (e.g., JSTOR). Whether such licenses extend to output-side copying remains contested, but they provide a good-faith defense and may satisfy EU AI Act Article 53(1)(c) documentation requirements.

6.4 Contractual Measures

User indemnification clauses. Commercial LLM terms of service should require users to indemnify the developer for any third-party claims arising from user-deployed outputs. Enforceability depends on relative bargaining power; business-to-consumer indemnity clauses may be deemed unfair under EU Directive 93/13/EEC. For enterprise customers, risk shifting via indemnification is standard and generally enforceable.

Limitation of liability clauses. Cap damages to the amount paid by the user (typically de minimis) and exclude consequential damages. Under the U.S. Uniform Commercial Code § 2-719, limitation-of-liability clauses are enforceable unless unconscionable. Under Chinese law, liability limitations that violate mandatory statutory provisions are void (Civil Code Article 497). Practically, contractual liability caps are less reliable in China compared with technical and organizational compliance measures.

Governing law and arbitration clauses. Parties may specify governing law favorable to the developer's liability position. Delaware law (U.S.) provides stable corporate liability jurisprudence. English law offers emerging judicial reasoning on generative AI liability. Chinese law imposes non-waivable mandatory administrative oversight that cannot be excluded by contract. For cross-border services, arbitration under ICC Rules in neutral venues such as Singapore or Geneva delivers high procedural predictability.

6.5 Integrated Strategy Matrix

Table 2 presents a decision matrix for legal avoidance strategies by jurisdiction and risk category.

Table 2 Legal Avoidance Strategy Matrix by Jurisdiction and Risk Category

Jurisdiction	Copyright Risk (Primary Strategy)	Privacy Risk (Primary Strategy)	Tort/Defamation Risk (Primary Strategy)
European Union	Technical: Opt-out respect, output filtering; Organizational: Conformity assessments	Technical: Differential privacy ($\varepsilon \leq 8$); Organizational: DPIA, ROPA	Technical: Output filtering for factual claims; Contractual: Liability caps (limited effect)
United States	Contractual: Indemnification, licensing; Technical: Output filtering (secondary)	Contractual: Terms of service; Technical: Prompt filtering for PII	Contractual: Indemnification, arbitration; Section 230 prompt defense
China	Technical: Output filtering (substantial similarity standard); Organizational: Algorithm filing	Organizational: PIPL compliance, consent mechanisms (challenging)	Technical: Truthfulness verification (limited feasibility); Administrative: CAC coordination

7 CONCLUSION

7.1 Summary of Findings

This paper has systematically analyzed platform-side infringement risks facing LLM developers engaged in commercial content generation, comparing the legal regimes of the European Union, United States, and China. The central finding is that traditional online intermediary liability frameworks—including EU DSA notice-and-action mechanisms, U.S. Section 230 immunity, and China's classic intermediary liability rules—are fundamentally ill-suited to generative AI systems. LLM developers are not passive information hosts but active content generators, and the stochastic, non-deterministic nature of LLM outputs makes ex ante filtering both technically challenging and legally indeterminate.
The three jurisdictions diverge significantly in their regulatory approaches. The EU imposes proactive due-diligence obligations via the AI Act's tiered risk framework, mandating documentation, transparency, and post-market monitoring. The U.S. relies on fragmented common law copyright and tort doctrines, while the applicability of Section 230 immunity to generative AI outputs remains unsettled, with courts and scholars divided. China enforces strict administrative oversight through algorithmic filing, content vetting, and direct liability for generative AI service providers.

7.2 Avoidance Strategy Synthesis

For LLM developers, the optimal avoidance strategy is hybrid: technical measures reduce the probability and magnitude of infringing outputs; organizational measures document compliance efforts and create audit trails; contractual measures allocate residual risk to users or counterparties.
The specific mix varies by jurisdiction. For EU operations, prioritize technical measures (output filtering, differential privacy) and documentation (conformity assessments) to demonstrate compliance with AI Act due-diligence requirements. Contractual indemnification is less effective because EU regulators target the developer directly regardless of private risk-allocation arrangements. For U.S. operations, prioritize contractual measures (indemnification, liability caps, arbitration) alongside minimal technical safeguards to support the legal defense that users, rather than the developer, function as primary content originators for prompt-directed outputs. For China operations, prioritize administrative compliance measures (algorithm filing, content labeling, CAC coordination) and output filters meeting the regulator's substantial similarity standard. Technical safeguards alone are insufficient in China, as regulatory requirements demand proactive content vetting rather than merely reactive filtering.

7.3 Limitations

Several limitations merit acknowledgment. First, legal frameworks across all three jurisdictions are rapidly evolving. As of mid-2026, multiple generative-AI-related cases are pending before courts, and new legislative proposals remain under consideration. This analysis reflects the regulatory and judicial landscape as of June 2026. Second, the study focuses on platform-side liability, with limited discussion of user-side or downstream distribution-chain liability. Third, empirical data on real-world enforcement patterns remains scarce due to the nascent nature of generative AI regulation. Fourth, the proposed compliance strategies have not yet been empirically validated, and their practical efficacy remains theoretical pending real-world deployment and testing.

7.4 Future Research Directions

Future research may examine quantitative patterns of cross-jurisdictional regulatory enforcement, testing whether specific technical compliance measures correlate with reduced regulatory sanctions and litigation risk. Controlled empirical experiments could quantify the effectiveness of varying output-filtering thresholds in mitigating substantially similar infringing outputs while preserving model performance. Furthermore, prospects for international regulatory

harmonization through treaties or model laws—potentially coordinated under the auspices of UNESCO or the G7—warrant further scholarly inquiry.

7.5 Concluding Remarks

The core normative question—whether LLM developers should bear primary liability for AI-generated content, or whether liability should instead fall on end users or downstream deployers—remains unresolved across major jurisdictions. In the absence of binding global legislative standards, a technically informed, jurisdiction-tailored compliance strategy represents the most pragmatic risk-mitigation approach. Developers that deploy robust technical safeguards, maintain thorough compliance documentation, and design contractual frameworks to allocate residual liability will be best positioned to navigate the increasingly fragmented global generative-AI legal environment.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCE

[1] IDC. Worldwide Artificial Intelligence Spending Guide. 2026. https://www.idc.com.
[2] State Cyberspace Administration of China. Interim Measures for the Administration of Generative Artificial Intelligence Services. 2023. https://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm.
[3] Lemley M A, Casey B. Fair learning. Texas Law Review, 2020, 99: 743. https://texaslawreview.org/fair-learning.
[4] Samuelson P. Generative AI meets copyright. Science, 2023, 381(6654): 158–161.
[5] National Development and Reform Commission, PRC. Reply to Proposal No. 4556 at the 3rd Session of the 14th National People's Congress. 2026. https://www.ndrc.gov.cn/xxgk/jianyitianfuwen.
[6] US Copyright Office. Copyright and artificial intelligence: a Report of the register of copyright (Part 1): Digital Replicas. US Government Publishing Office, 2024.
[7] US Copyright Office. Copyright and artificial intelligence: a Report of the register of copyright (Part 2): Copyrightability. US Government Publishing Office, 2025.
[8] Nasr M, Carlini N, Jagielski M, et al. SCALPEL: Exploring the Limits of Extraction Attacks on LLMs with Fine-tuning. Proceedings of the 2023 IEEE Symposium on Security and Privacy (S&P), 2023.
[9] Court of Justice of the European Union. YouTube LLC and Cyando AG v. Frank Peterson and Google Germany GmbH (Case C-682/18). ECLI:EU:C:2021:503, 2021.
[10] European Commission & European Board for Digital Services. First report on the most prominent and recurrent systemic risks on very large online platforms and very large online search engines under the Digital Services Act. Publications Office of the European Union, 2025.
[11] European Data Protection Board. Guidelines 2/2024 on processing personal data for training generative AI models. EDPB Document, 2024.
[12] Russinovich M, Salem A. Obliviate: Efficient Unmemorization for Protecting Intellectual Property in Large Language Models. 2025. https://ar5iv.labs.arxiv.org/html/2502.15010.
[13] Zhang J, Yu J, Marone M, et al. Certified Mitigation of Worst-Case LLM Copyright Infringement (BloomScrub). 2025. https://arxiv.org/abs/2504.16046.
[14] Abadi M, Chu A, Goodfellow I, et al. Deep learning with differential privacy. Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, 2016: 308–318. https://doi.org/10.1145/2976749.2978318
[15] Kirchenbauer J, Geiping J, Wen Y, et al. A watermark for large language models. Proceedings of the 40th International Conference on Machine Learning, 2023: 17061–17084.