

Comments of the International Center for Law & Economics

*RE: Proposed Rule 15 CSR 60-19.020 Prohibition on
Restricting Choice of Content Moderator*

July 14, 2025

Authored by:

Kristian Stout (Director of Innovation Policy, International Center for Law & Economics)

Ben Sperry (Senior Scholar, Innovation Policy, International Center for Law & Economics)

Executive Summary

The International Center for Law & Economics (ICLE) respectfully submits these comments in opposition to proposed rule 15 CSR 60-19.020, which would mandate "choice screens" for content moderators on social media platforms. This rule is economically inefficient, technically unworkable, and legally vulnerable, representing a regulatory solution that is demonstrably worse than the problems it purports to address.

The proposed rule rests on four fundamentally flawed premises that render it unsuitable for implementation. First, the rule creates systematic privacy and security vulnerabilities by mandating an architecture functionally identical to systems that enabled major data breaches, including Cambridge Analytica. The requirement for broad API access to third-party moderators multiplies attack surfaces, ignores the "other people's data" problem where individual user choices expose non-consenting network connections to unknown third parties, and dangerously fragments responsibility for protecting vulnerable populations across an ecosystem of unvetted entities.

Second, the rule's central mechanism—the choice screen—ignores over a decade of documented failures from European Union regulatory experiments. Browser choice screens and search engine choice screens have consistently failed to alter market dynamics, with research demonstrating that users reveal preferences for streamlined, familiar experiences over additional complexity.

Third, the rule reflects a profound misunderstanding of content moderation as a technical system. Content moderation is not a simple "plug-and-play" module but rather a deeply integrated, multi-layered system inseparable from platform architecture and business models. The mandate for "interoperable access to data" would fragment conversational context, create unresolvable conflicts between competing moderation systems, and erect prohibitive barriers to entry that would likely recreate the same market concentration the rule seeks to address.

Fourth, the rule faces serious constitutional and federal-preemption challenges. The mandate likely violates the First Amendment's compelled speech doctrine by forcing platforms to carry content against their editorial judgment, contradicting the U.S. Supreme Court's holding in *Moody v. NetChoice* that content moderation decisions are protected "expressive products." The rule also conflicts with Section 230's federal protections for content moderation.

The Missouri Attorney General should withdraw this proposed rule entirely. If regulation of social media platforms is desired, policymakers should focus on transparency requirements and user control mechanisms rather than mandating technically unworkable architectures that compromise user privacy and security while failing to achieve their stated objectives.

I. Privacy and Security Risks

The most dangerous aspect of the proposed rule is not its economic inefficiency or technical infeasibility, but its disregard for the privacy and security of Missouri's citizens. The mandate for

“interoperable access to data” creates a blueprint for privacy and security incidents by forcing platforms to construct an architecture that is inherently insecure.

The technology industry has learned important lessons about API security and data access from past incidents, most notably the Cambridge Analytica episode. That incident involved data from over 50 million Facebook profiles being accessed through Facebook's Graph API v1.0, which allowed third-party applications to access not only data from users who installed apps, but also data from their entire network of friends without those friends' knowledge or consent.¹

The Missouri rule's requirement for “interoperable access to data...for the purpose of moderating what content is viewed by the user” presents similar architectural challenges. For a third-party service to moderate content in a user's feed, that service would necessarily need access to content posted by the user's friends, connections, and followed accounts. There is likely no technical way to provide the content moderation service required by the rule without also providing access to this network-level data.

The proposed rule appears to disregard these lessons by requiring platforms to create similar broad data access mechanisms for third-party moderators. Rather than building on the industry's improved understanding of API security risks, the rule would require platforms to recreate data access patterns that have proven problematic in the past.

The risk of creating powerful APIs for third-party access is not restricted to just a few high-profile incidents—it is documented by a consistent pattern of major data breaches caused by insecure API implementation or malicious abuse. Parler had its entire database of user posts and information exposed because its public-facing API lacked proper authentication.² Clubhouse suffered a similar breach exposing data from 1.3 million users due to the same vulnerability.³ A flaw in Twitter's API allowed attackers to link email addresses and phone numbers to 5.4 million user accounts, which were then sold online.⁴

These incidents demonstrate that forcing the creation of broad API access for multiple third parties exponentially increases the “attack surface” of platforms, making data breaches not just possible but inevitable. Each additional third-party moderator represents another potential point of failure,

¹ See, e.g., Gus Hurwitz, *Soylent Analytica: The Graph Is Too Damn Open*, TRUTH ON THE MARKET (Mar. 21, 2018), <https://truthonthemarket.com/2018/03/21/soylent-analytica-the-graph-is-too-damn-open>; Nathaniel Fruchter, Michael Specter, & Ben Yuan, *Facebook/Cambridge Analytica: Privacy Lessons and a Way Forward*, INTERNET POL'Y RESEARCH INITIATIVE (Mar. 20, 2018), <https://internetpolicy.mit.edu/blog-2018-fb-cambridgeanalytica>.

² See, e.g., Ran Ilany, *5 Real-World API Security Breaches from 2021*, OUTSHIFT (last updated Jun. 17, 2025), <https://outshift.cisco.com/blog/real-world-api-security>.

³ *Id.*

⁴ See Jan Cornet, *The True Cost of API Security Breaches: Examples, Consequences & Prevention*, SEEBURGER BLOG (Mar. 6, 2025), <https://blog.seeburger.com/the-true-cost-of-api-security-breaches-examples-consequences-prevention>.

another set of credentials that could be compromised, and another organization with varying levels of security sophistication in handling sensitive user data.

Further, the rule is built on a fundamental fallacy—the idea that consent in choosing a moderator is purely individual. It ignores the reality of networked privacy, where one person's choice has direct consequences for many others' data. When User A chooses a third-party moderator, they are not just consenting for themselves—they are unilaterally granting an unknown entity access to posts, photos, and personal information shared by Users B, C, and D, none of whom consented to this new party accessing their data.

The proposed rule multiplies attack surfaces by requiring platforms to expose sensitive data to multiple third-party moderators of varying security sophistication. While major platforms invest heavily in cybersecurity, third-party moderators may lack comparable resources or expertise, creating weak links in the data security chain. Each additional moderator represents not just another potential breach point, but another organization that could be targeted by sophisticated attackers seeking access to platform data.

Thus, the proposed rule transforms what should be a controlled, secure environment into a distributed system with multiple points of failure, making large-scale privacy breaches not just likely but inevitable. The state would bear responsibility for bringing these vulnerabilities into existence, putting Missouri citizens' personal data at systematic risk through regulatory mandate.

Beyond the direct privacy and security vulnerabilities, the proposed rule creates a dangerous diffusion of responsibility for protecting vulnerable populations, particularly children and victims of violence. Currently, platforms maintain centralized systems for identifying and responding to serious threats like child sexual abuse material (CSAM), terrorism, and imminent violence. These systems include mandatory reporting channels to the National Center for Missing & Exploited Children, direct collaboration with law enforcement agencies, and escalation procedures for time-sensitive threats.⁵

The proposed rule could fragment these critical safety functions across an ecosystem of third-party moderators with varying levels of expertise, resources, and commitment to user safety. Most concerning, the rule contains no requirements that third-party moderators maintain the same reporting obligations, technical capabilities, or law enforcement coordination that major platforms currently provide. A permissive third-party moderator chosen by a user could lack the sophisticated detection systems, trained personnel, and established protocols necessary to identify CSAM or credible threats of violence.

⁵ See, e.g., *Online Child Protection*, META SAFETY CENTER (last accessed Jul. 7, 2025), <https://about.meta.com/actions/safety/onlinechildprotection>; *Transparency Report: Jul. 1, 2024-Dec. 31, 2024*, SNAP VALUES (Jun. 20, 2025), <https://values.snap.com/privacy/transparency>.

This creates a digital version of the bystander effect, where responsibility for protecting vulnerable users becomes so diffused across multiple parties that critical threats fall through the cracks.⁶ When a concerning post could theoretically be handled by the platform, a third-party moderator, or neither (depending on user choices and system conflicts), the likelihood of appropriate intervention is likely to decrease because of ambiguity in control over the situation. Unlike individual bystander scenarios, this regulatory structure institutionalizes diffused responsibility at scale, potentially affecting millions of users.

The rule provides no mechanism to ensure that third-party moderators possess the technical infrastructure to detect sophisticated threats, the trained personnel to evaluate context-dependent dangers, or the established relationships with law enforcement necessary for rapid response. By mandating that platforms “must not override the content moderation decisions of competing content moderators,” the rule could actively prevent platforms from protecting users when third-party moderators fail to identify serious threats.

II. Choice Screens Have Been Shown Not to Work and to Frustrate Users

The proposed rule's central mechanism—the “choice screen”—rests on a fundamentally flawed assumption that simply presenting users with options will meaningfully alter market dynamics. Over a decade of regulatory experiments in the European Union have demonstrated that this assumption is wrong.

A. The Reality of User Preferences and Behavior

A fundamental problem with choice screens lies in their failure to account for revealed consumer preferences. Research conducted by Mozilla into browser choice interventions confirms that users consistently demonstrate through their actions that they prefer streamlined, frictionless experiences over additional decision-making burdens, with consumers regularly ignoring or dismissing pop-ups and banners that interrupt their intended workflow.⁷

When users are focused on tasks like setting up new accounts, their revealed preferences consistently show they want to complete these tasks efficiently (or not at all).⁸ Prompts to make complex decisions about “content moderators”—a concept most users neither understand nor actively seek—will be met with the rational consumer response of dismissal in favor of proceeding with their primary objective.

⁶ See, e.g., Ruud Hortensius & Beatrice de Gelder, *From Empathy to Apathy: The Bystander Effect Revisited*, 27 CURR. DIR. PSYCHOLOGICAL SCI. 249 (2018), available at <https://pmc.ncbi.nlm.nih.gov/articles/PMC6099971> (discussing the bystander effect and possible neural mechanisms for its occurrence).

⁷ Gemma Petire, *Beyond Choice Screens: Exploring browser choice design interventions*, MOZILLA RESEARCH (last accessed Jul. 7, 2025), <https://research.mozilla.org/browser-competition/remedyconcepts> (Notably, even Mozilla's research, which advocates for choice screen interventions as part of its business interest, acknowledges that user behavior consistently demonstrates preferences for stable, familiar systems over additional complexity).

⁸ *Id.*

Users' consistent selection of pre-installed, familiar defaults reveals their preference for systems that work without requiring additional cognitive effort or technical understanding.

Further, this pattern of user behavior reflects rational consumer choice rather than psychological bias. Users demonstrate through their actions that they value convenience, familiarity, and reduced complexity. The widespread persistence of default settings across technology platforms reflects genuine consumer preferences for systems that function without requiring additional decisions from non-expert users who lack both the technical knowledge and the incentive to research content moderation alternatives.

B. EU Evidence: A History of Ineffectiveness

The most telling evidence against choice screens comes from the European Union's repeated failures to achieve meaningful market changes despite multiple iterations and refinements.⁹ The browser choice screen imposed on Microsoft following antitrust enforcement had such negligible impact that the screen was defunct for months due to a software bug—a lapse that went unnoticed by both regulators and the market, demonstrating the profound lack of user engagement with the remedy.¹⁰

More recently, the EU's mandate for an Android search engine choice screen has proven similarly ineffective.¹¹ Research shows that the mandate had no lasting effect, with the few users who selected alternatives often reverting to the default search engine after a short period.¹²

Even where modest gains have been reported under the Digital Markets Act (“DMA”) framework, these cannot be attributed to choice screens alone. The DMA imposes a complex web of obligations including interoperability mandates, prohibitions on self-preferencing, and detailed compliance reporting—a far more complex regulatory regime than the proposed singular rule. To suggest that the proposed rule could replicate the contested and at best limited success of the EU's comprehensive framework is illogical.

III. Integration Reality: The ‘Plug-and-Play’ Fallacy

The proposed rule is premised on a fundamental misunderstanding of how content moderation actually works within modern digital platforms. The proposed rule treats moderation as a simple, detachable “plug-and-play” module that users can swap out like changing a phone case. This view is

⁹ See Geoffrey Manne & Dirk Auer, *Antitrust Dystopia and Antitrust Nostalgia: Alarmist Theories of Harm in Digital Markets and Their Origins*, 28 GEORGE MASON L. REV. 1279, 1385-89 (discussing the limitations and lack of success from the Microsoft remedies).

¹⁰ *Id.* at 1385 (“[T]he browser choice screen remedy was so ineffective [that, when Microsoft illegally stopped implementing it, it took authorities and consumers a full fourteen months to notice.]”).

¹¹ See, e.g., Dirk Auer, *The Future of the DMA: Judge Dredd or Juror 8?*, TRUTH ON THE MARKET (Apr. 8, 2024), <https://truthonthemarket.com/2024/04/08/the-future-of-the-dma-judge-dredd-or-juror-8>.

¹² See Spence Purnell, *One of the Flaws in DOJ's Anti-Trust Case: People Overwhelmingly Choose Google*, REASON FOUNDATION (Oct. 4, 2023), <https://reason.org/commentary/one-of-the-flaws-in-doj-s-anti-trust-case-people-overwhelmingly-choose-google>.

not only technically incorrect but reveals a profound misunderstanding of the integrated, complex nature of content moderation systems that has developed over decades of platform evolution.¹³

Content moderation is not a simple filter applied at the end of a content pipeline—it is a deeply integrated, multi-layered system inseparable from a platform's core architecture and business model.¹⁴ Effective moderation relies on a complex “stack” of technologies and human processes working in concert: proactive AI-driven systems that scan vast volumes of content in real-time; reactive systems that process and prioritize user reports; detailed decision trees and guidelines for human reviewers; multi-tiered escalation paths for nuanced cases; and extensive quality assurance loops to ensure consistency.¹⁵

This entire apparatus is woven into the fabric of how content is ingested, analyzed, ranked, and displayed. The platform's economic incentives—whether advertising-based models that prioritize engagement or subscription models focused on user utility—directly shape investment in and strategy for content moderation. A platform's recommendation algorithms, user interface design, and data collection practices are all interconnected with its moderation systems.¹⁶

To mandate that this intricate, integrated system be opened to external third-party providers fundamentally misunderstands how content moderation actually works. It would be like requiring an airport to allow third parties to take over air traffic control for individual flights while expecting the same level of safety and coordination—the real-time, interconnected nature of the system makes such fragmented control impossible.

Further, the proposed rule's demand for “interoperable access to data”¹⁷ would force platforms to create systems analogous to decentralized social networks like Mastodon, where independent servers attempt to moderate fragmented conversations. Research into moderation in decentralized environments reveals a critical flaw: fragmented conversational context.¹⁸

¹³ See Giovanni Sartor & Andrea Loreggia, *The Impact of Algorithms for Online Content Filtering or Moderation*, at 20-23 (Study requested by the JURI committee of European Parliament), available at [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/657101/IPOL_STU\(2020\)657101_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/657101/IPOL_STU(2020)657101_EN.pdf).

¹⁴ *Id.* at 23 (noting that “[i]n big platforms filtering is entrusted to a complex socio-technical system, including a multi-stage combination of human and machines that interact in complex ways”).

¹⁵ See *id.* at 36-44.

¹⁶ To take one example, Meta maintains a closed advertising ecosystem across all of its products that allow it to tightly integrate ad experiences with the content a user interacts with and prefers. By design, it is a highly integrated experience, including the content moderation services it provides. See, e.g., *Audience ad targeting*, Meta (last accessed Jul. 7 2025), <https://www.facebook.com/business/ads/ad-targeting>. Changing one element of this formula by necessity affects all the other aspects of it – including the ability of Meta to offer the services to users at the current price and quality levels.

¹⁷ 15 CSR 60-19.020(2)(C).

¹⁸ Vibhor Agarwal et al., *Decentralized Moderation for Interoperable Social Networks: A Conversation-Based Approach for Pleroma and the Fediverse*, 18 PROCEEDINGS OF THE INTERNATIONAL AAAI CONFERENCE ON WEB & SOCIAL MEDIA 2, 3-4.

In the proposed rule's envisioned system, a single conversation could be fragmented across multiple moderation services. Depending on implementation, a third-party moderator might have only partial visibility into interactions—seeing a reply without the original post, or a comment without the full thread of preceding remarks. Crucially, third-party moderators would likely lack access to the full network graph, user history, and behavioral signals essential for interpreting meaning and context.

Without this context, accurate moderation decisions about nuanced forms of speech—sarcasm, inside jokes, political satire, or coded harassment—become impossible. A third-party moderator operating under this rule would be making decisions on isolated content fragments, stripped of the context necessary for accurate interpretation. This would inevitably lead to both false positives (censorship of benign speech) and false negatives (failure to remove genuinely harmful content).

Moreover, the proposed rule creates an unworkable conflict between platform moderation and third-party decisions that has no technical or legal resolution mechanism. Paragraph (6) of the rule allows platforms to moderate certain content categories like child sexual abuse material or incitement to violence, irrespective of third-party moderator choices. But the rule provides no framework for resolving conflicts when these two systems reach different conclusions about the same content.

If a user selects a permissive third-party moderator, but the platform's systems flag content under its “good-faith judgment” as inciting violence, which decision prevails? The rule is silent on this critical question, guaranteeing legal uncertainty and technical chaos as two separate moderation systems attempt to operate simultaneously on the same content stream. This creates an impossible situation where platforms must simultaneously comply with conflicting mandates—allowing content per the third-party moderator while removing it per their own safety obligations.

IV. Constitutional and Legal Issues

Beyond its practical challenges, the Missouri rule faces fundamental constitutional and statutory conflicts that render it legally vulnerable. The rule's mandates directly implicate platforms' First Amendment rights and conflict with established federal protections for content moderation.

A. First Amendment: Compelled Speech Violations

The U.S. Supreme Court has repeatedly affirmed that content moderation constitutes protected editorial discretion under the First Amendment. In *Moody v. NetChoice*, the Court explicitly stated that when platforms use their standards to decide which content to display or how to organize it, “they are making expressive choices” that “receive First Amendment protection.”¹⁹ This protection is rooted in the longstanding principle from *Miami Herald Publishing Co. v. Tornillo* that the

¹⁹ *Moody v. NetChoice LLC*, 144 S.Ct. 2382, 2406 (2024).

government cannot compel private actors to host or disseminate speech they would prefer to exclude.²⁰

The proposed rule likely runs afoul of this established doctrine by mandating that platforms must not “override the content moderation decisions of competing content moderators.”²¹ This forces platforms to carry and display content that may violate their own terms of service, community standards, and expressive goals. A platform prioritizing family-friendly content, for example, could be compelled to display graphic material if a user unintentionally selects a third-party moderator with more permissive standards. This represents classic compelled speech—forcing a private entity to serve as a conduit for messages it finds objectionable.

As the Supreme Court clarified in *Moody*, a state “may not interfere with private actors' speech to advance its own vision of ideological balance.”²² The proposed rule likely violates this principle by mandating the substitution of third-party editorial judgment for platforms' own constitutionally protected right to exercise editorial discretion.

B. Section 230 Preemption: Federal Conflict

The proposed rule also risks conflicting with federal statutory law, specifically Section 230(c)(2)(A) of the Communications Decency Act. This “Good Samaritan” provision explicitly protects platforms from liability for “any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected.”²³

The Missouri rule directly undermines this federal protection by seeking to impose state-law liability on platforms precisely for exercising their federally protected content moderation judgment when it conflicts with user-chosen third-party moderators. The rule effectively punishes platforms for actions that Section 230(c)(2)(A) is designed to immunize.

Federal law generally preempts conflicting state laws,²⁴ and Section 230(e)(3) explicitly codifies this principle: “No cause of action may be brought and no liability may be imposed under any State or local law that is inconsistent with this section.”²⁵ A platform cannot simultaneously comply with the proposed rule's mandate to defer to third-party moderators and exercise its federally protected right to moderate content according to its own good-faith judgment. This direct conflict makes the rule

²⁰ See *id.* at 2400 (“The seminal case is *Miami Herald Publishing Co. v. Tornillo*, 418 U.S. 241 (1974)... the cure proposed, it concluded, collided with the First Amendment's antipathy to *state* manipulation of the speech market.”).

²¹ 15 CSR 60-19.020(5).

²² *Moody*, 144 S.Ct. at 2407.

²³ 47 U.S.C. §230(c)(2)(A).

²⁴ See, generally, Bryan L. Adkins, Alexander H. Pepper, & Jay B. Sykes, *Federal Preemption: A Legal Primer*, CRS Report R45825 (May 18, 2023), available at <https://www.congress.gov/crs-product/R45825>.

²⁵ 47 U.S.C. §230(e)(3).

likely preempted by federal law, creating additional legal uncertainty for any platform attempting compliance.

Conclusion

The proposed rule 15 CSR 60-19.020 fails any reasonable cost-benefit analysis. The overwhelming evidence demonstrates that the rule creates unacceptable privacy and security risks, choice screens do not work, the technical architecture mandated by the rule is unworkable, and the constitutional challenges are likely insurmountable. The rule would impose substantial costs on platforms and users while failing to achieve its stated objectives of promoting competition or consumer choice in content moderation.

The costs are clear and significant: systematic security vulnerabilities that put Missouri citizens' personal data at risk and fragment protection for vulnerable populations; millions of dollars in API development expenses that favor large corporate moderators over innovative alternatives; degraded user experiences through fragmented moderation systems; and legal uncertainty from constitutional and federal preemption challenges. Against these substantial costs, the rule offers no credible benefits—choice screens have repeatedly failed in other contexts, and the technical requirements would likely recreate the same market concentration the rule purports to address.

The Missouri Attorney General should withdraw this proposed rule entirely. The rule represents an expensive regulatory intervention that would ultimately fail to achieve its objectives while imposing substantial costs and security risks on Missouri citizens, particularly the most vulnerable users who depend on effective content moderation systems for their protection.

If the Attorney General believes social media regulation is necessary, far better approaches exist that do not mandate unworkable technical architectures or compromise user security. Effective alternatives might include transparency requirements that allow users to understand how content moderation decisions are made, user control mechanisms that give individuals more granular choices about their own content consumption, or disclosure requirements that help users make informed decisions about platform use.

Such approaches would respect the technical realities of content moderation systems, avoid creating systematic security vulnerabilities, and preserve platforms' constitutional rights while potentially achieving legitimate regulatory objectives. Most importantly, these alternatives would not require platforms to reconstruct their fundamental architectures in ways that have proven both ineffective and dangerous.