4-2024

Content Moderation and the Least Cost Avoider

Paul Rosenzweig

Content Moderation and the Least Cost Avoider

Paul Rosenzweig*

Who is responsible for mitigating the harm from malicious social media content? The poster? The social media platform? The ISP that carries the transmission? Or some other actor? In this article, I argue that we can learn a useful lesson from the economic concept of the "least cost avoider" and that, properly understood, the concept suggests that significant responsibility for reducing harmful content should be allocated to non-platform actors.

Introduction

Today, Section 230 of the Communications Decency Act effectively protects social media platforms from liability for the content that their users post. If, for example, I post a splenetic antisemitic screed on Twitter (a/k/a "X") advocating violence against Jewish students, I might bear personal responsibility for posting its content, but Twitter/X does not bear any for hosting my diatribe.

Many find this state of affairs troubling. Some of those, for example, who wish to reduce the dissemination of Child Sexual Abuse Material ("CSAM") think that one way to achieve this objective would be to make social media platforms that host CSAM liable for damages that might arise from providing a platform for that content. And there is some justification for that instinct – liability is a deterrent to conduct and if we do not want CSAM proliferating in the online information ecosystem we might reasonably think that assigning blame for hosting that content to those who control the platforms would be a reasonable way to deter them from continuing to do so.

But the idea of liability is distinct from the question of ***who*** should bear liability. Even if we think that someone in the internet ecosystem should bear the responsibility for managing content and removing malicious or harmful information, it is not necessarily obvious that the "someone" in question should be the social media platforms. It could readily be any number of other actors in the system, ranging from ISPs to security service providers and anyone in between.

An answer is urgently needed, as the problem of malicious content is only going to get worse. On one hand, the advent of artificial intelligence portends an ever-greater challenge from the creation of realistic deep fakes. Add to this the fact that recent corporate changes (most notably at Twitter/X) reflect a growing unwillingness of social media platforms to serve as gatekeepers of content as they return to their original model of generally unregulated publication of information. Taken together, these changes threaten an explosion of online harm.

* This project received funding support from the Anti-Defamation League through the Tech, Law & Security Program at the Washington College of Law, American University. TLS maintains strict intellectual independence and sole editorial direction and control over its intellectual property, ideas, projects, publications, events, and other research activities. Consistent with TLS policy, the content of this essay reflects the views of its author alone.

Yet mitigating online harm is not costless. Any degradation in the freedom to speak raises the possibility of over-limitation. It likewise offers the significant possibility of error, as non-harmful content is mistakenly lumped with harmful materials.

How shall we mitigate online harms most effectively while remaining cognizant of the countervailing values of free expression and privacy? This paper suggests an approach to a solution to the problem. It suggests that content moderation should borrow from existing economic theory and impose obligations on the "least cost avoider" – that is on the provider who is best able to ameliorate the harm at the lowest social cost (where cost is understood as both monetary and hedonic).

The analysis proceeds in three parts. First, it begins with a brief outline of the underlying economic concept of a least cost avoider (sometimes also called a "cheapest cost avoider"). Then the paper, building on the earlier work of Ruddock and Sherman, briefly outlines the online information ecosystem identifying possible entry points for content moderation and control. Finally, the paper turns to a detailed analysis of the costs and benefits of those various entry points as avenues for moderation and control.

In the end, the paper concludes that beyond social media platforms there are other entities within the internet eco-system (such as search engine providers and web-hosting services) who are highly plausible entry points for content moderation and control. As we seek to mitigate the harms from malicious content we should broaden the lens of our approach and consider a multi-pronged approach that recognizes the complexity of the internet environment.

The Least Cost Avoider

In any given social situation where the potential for harm exists, there will be multiple actors who might be capable of taking steps to avoid the potential harm in question. In a car accident, there are typically two drivers, either of whom might have a chance to avoid the harm. In the construction of a ladder with a flaw in it there are multiple actors ranging from the designer to the factory builders, to the dealer who sold the ladder, and possibly even the purchaser, if they modified the ladder in some way. It has long been understood that this reality is equally true for deliberate harms like pollution – the possible ways to avoid the harm vary across many dimensions. Though we don't often think of it this way one possible mitigation for, say, toxic air pollution would be the provision of gas masks. Not, obviously, a good social solution – and not one that any reasonable person would advocate – but a good illustration of the reality that harms are multivariate in nature.

The question then that economists ask is who among these many potential actors is the one best placed to minimize the costs arising from potential harm. Those costs are threefold. They include: the costs of the injury were it to occur; the costs involved in avoiding the injury and preventing the harm; and the administrative costs associated with allocating the responsibility in the first place and adjudicating it if the responsibility is in dispute.

This actor, whomever it might be, is known as the least cost-avoider, or sometimes the cheapest avoider. Avoiding harm is not costless, but the goal is to identify the actor who can minimize those costs as best as possible – in other words best placed to minimize the sum of the three types of cost. As Guido Calabresi, the noted Yale legal scholar, put it: "[T]he search for the cheapest avoider of accident costs is the search for that activity which has most readily available a substitute activity that is substantially safer.

It is a search for that degree of alteration or reduction in activities which will bring about primary accident cost reduction most cheaply."

Identifying this cheapest cost avoider can be difficult. "[T]he chosen loss bearer must have better knowledge of the risks involved and of ways of avoiding them than alternate bearers; he must be in a better position to use that knowledge efficiently to choose the cheaper alternative; and finally he must be better placed to induce modifications in the behavior of others where such modification is the cheapest way to reduce the sum of accident and safety costs. The party who in practice best combines these not infrequently divergent attributes is the 'cheapest cost avoider' of an accident who would be held responsible for the accident costs under the market deterrence standard."

It may seem strange to think of content moderation as equivalent to accident avoidance, but fundamentally the problems are identical in conception. In each case a harm may exist; there may be costs to avoiding the harm (including, for example, social costs from prohibiting free speech); and in each case, there are a multiplicity of actors who might reasonably be thought of as in a position to take action to minimize the costs arising from the harm.

The Information Ecosystem

In their 2020 paper, "*Widening the Lens on Content Moderation*," two colleagues from the Tech, Law & Security Program, Jenna Ruddock and Justin Sherman, described the nature of the "online information ecosystem" – an ecosystem that is much broader than the well-known social platforms that host content (such as Twitter/X or Facebook). As they recounted in greater detail, this broader ecosystem includes several possible actors who might be assigned liability and/or responsibility for mitigating harmful content.

We might reasonably characterize the broader set of responsible providers as follows: First, there are the logical services of the network. That is those services that are necessary for accessing, browsing, delivering, hosting, and securing information online. In thinking about these services we can identify several whose activities are essential:

- Internet Service Providers (ISPs) who afford individuals and enterprises access to the network through direct Internet connections;
- Virtual Private Network (VPN) operators who provide an alternate means of accessing the network by creating a private connection through a public entry point;
- Domain Name System (DNS) operators, including registrars and registries who maintain the hierarchical and distributed internet naming system for computers and services and allow navigation of the network by associating information with domain names assigned to each entity;
- content delivery networks (CDNs) like Cloudflare and Akamai who operate geographically distributed groups of servers that cache content close to end users, thereby permitting the quicker transfer of assets needed for loading Internet content, such as HTML pages, JavaScript files, stylesheets, images, and videos;
- cloud service providers (like AWS) who allow users to store data and access services in a distributed way;
- web hosting platforms like Bluehost that offer users a facility to create and maintain a website;

- DDOS mitigation services that protect hosts and servers by filtering attacks on the system; and
- web browser systems, like Chrome or Safari that provide the digital gateway to access content on the public internet (and also the dark web).

Second, there are a group of services that work directly with content, in some ways aggregating or curating that content. This category includes well-known social platforms (such as X/Twitter or Facebook or TikTok) that typically are looked to for content moderation. But it also would include, for example, online marketplaces, such as Amazon or Alibaba which directly curate content and rank and present it to users.

Similarly, this category would include search engines (like Google Search or DuckDuckGo) that algorithmically curate search results to specific queries and present them to users. Increasingly, we can add to this curating function artificial intelligence sorting systems like ChatGPT which respond to queries by, in effect, curating and collating responsive information. And here we might also include app stores (like Google Play and Apple's App Store) that allow users to download applications to their devices and have both technical and content-oriented rules governing which applications are made available.

Finally, the ecosystem also includes a group of financial services that facilitate monetary exchange (like PayPal or Stripe). These services lie at the core of much of the information exchange relating to e-commerce or fee-for-service online interactions.

Of course, we should acknowledge that many providers offer multiple services and that the lines between different types of offerings can tend to blur. That having been said, one can easily discern that multiple providers are responsible for administering various aspects of information exchange along the network. Seen in that light, one should consider the possibility of assigning responsibility for content moderation more broadly than by simply looking to the social media platforms to mitigate the harm.

<u>The Concept of Cost in the Information Eco-System</u>

Who is best situated to mitigate the risks of harmful content with minimal societal cost? The final piece of the puzzle is to have a better understanding of the idea of "cost" which is often broader than the narrow concept of financial cost.

On the face of it, the information eco-system poses a classic economic question of who in a long supply chain of conduct is the least or cheapest cost avoider. However, the analysis must begin with the recognition of a significant (perhaps even predominant) difference between traditional least-cost analysis and the situation presented. In most typical situations the "costs" involved in accident or incident avoidance are predominantly (or even exclusively) internal to the cost-avoiding enterprise. Ford needs to add additional protection to its gas tank or McDonalds needs to exercise better care to keep its meat frozen until used. Meanwhile, the benefits from the cost expenditure are typically public in nature – safer cars and safer meat for all.

Thus, the least-cost analysis is often thought of as a way of mitigating natural externalities through the imposition of legal obligations. But here – in the case of social media – we must account for the fact that both the costs and benefits of action (or inaction) will involve costs to the general public. In one version, we mitigate malicious content; in another version, we over-correct and limit useful, beneficial public discourse.

The underlying theory has long been understood: We think of social media as a private good. In many instances, (including this one) the production of a private good will cause an externality – that is, the activity between two economic actors may directly and unintentionally modify a third party's utility. Externalities can be either positive (as when a transaction I voluntarily enter into benefits a third party who pays nothing for the benefit) or negative (when the transaction harms an individual).

Many social media activities have positive externalities. For example, by enabling communications amongst all users, social media systems benefit others on the network who are derivatively made better informed. Indeed, as a general matter, the fundamental premise of a free society is that interactions raise the level of knowledge and discourse and are thus to be fostered.

But social media also has negative externalities. The harm caused by some discourse is not fully captured in the costs of operating the information ecosystem. Economists call this a pricing problem: private sector actors often do not internalize the costs of malicious content failures in a way that leads them to take adequate protective steps. When content moderation fails to prevent the distribution of malicious content or when a provider fails to interdict a disinformation attack, there is no mechanism through which to hold the any member of the information ecosystem responsible for the costs of those failures. Consequently, the costs are borne entirely by the end users. In this way, social media harms on the broader Internet are a classic market externality, the true costs of which are not adequately recognized in the prices charged and costs experienced by individual or corporate actors.

Least-cost analysis thus asks the question of how best to change that incentive structure by putting a regulatory obligation on the actor who is most likely able to mitigate the harm with the least cost. In this case, of course, the idea of cost includes both actual financial costs (of say operating a content moderation system) and the external costs that might arise from the false positive moderation of content that is or could be societally beneficial.

[Though beyond the scope of this paper, it is worth noting that regulation is only one of several possible options for government intervention – others include the possibilities of subsidizing good behavior; taxing bad behavior; or allowing the imposition of civil liability. For purposes of this analysis, the author is agnostic as to which form of coercive intervention might be best – especially since they are often economically equivalent.]

The Information Ecosystem Least Cost Avoider

One cannot do a quantitative analysis of the costs that arise in the information ecosystem. The data either does not exist at this time or, to the extent that it may, it is not readily publicly available. But that does not disable us from conducting a qualitative analysis of the least-cost avoider question. As one might expect, the results are mixed – not all factors point in one direction. But at the same time, the import of the analysis is clear – the scope and ambit of content moderation is broader than our current focus on social media platforms suggests. Good policy requires broadening the lens.

We may assess the information environment along five different avenues of inquiry:

- Which parts of the ecosystem have better knowledge of the risks involved?
- Which have better ways of avoiding the risks/harms than alternate bearers?

- Which are in a better position to use that knowledge efficiently to choose the cheaper alternative?
- Which, by acting, will impose the least negative social costs through the risk of over-limitation?
- Which are better placed to induce modifications in the behavior of others where such modification is the cheapest way to reduce the sum of all social costs?
- And, related to these last two which are better positioned to fine-tune the moderation processes and which are blunter instruments?

As we have already noted, the answer to these questions is complex and confounded. Some parts of the information eco-system, like the social media platforms, operate on a retail basis – at the level of individual messages – and they thus have greater granular knowledge of malicious content but less systematic capacity to effect change. By contrast, for example, CDNs operate at a wholesale level – they can be aware at a gross level of the nature of the content involved and they can have a greater and more pervasive impact when they act. At an even higher level, ISPs have plenary visibility into the content on their networks but they are often constrained by law from acting and, when they do act, they do so at a very high level of control with broad-sweeping impact.

The following chart attempts to qualitatively map the various distinctions amongst the multitude of actors within the ecosystem, assigning a blue/yellow/orange/red color code to reflect an assessment of how well, or poorly, a particular actor in the system is capable of impacting the flow of malicious content in the context of various capability questions:

	Knowledge of Risks	Better Risk Avoidance	Efficiency/Low Cost	Least Negative Social Cost and Bluntness of Tool	Most Coercive/Capable of Inducing Modification
ISPs	Low visibility into underlying content. ISPs do not normally review content on their networks.	If authorized to do so ISPs would be able to readily track and review content. Would require legal change, however, and is therefore problematic.	Exceedingly efficient. ISP access is a chokepoint for content distribution. Restriction here is relatively easy to achieve.	High risk of negative social costs. Denial of ISP access is complete denial of ability to produce and distribute content. Unlike other possible tools, ISP access is not a granular or precise tool.	Highly coercive of change. Denial of ISP access would significantly impact user, compelling modification of behavior. Not easy to avoid sanction.
VPNs	Low visibility into underlying content. VPN services will not have access to content.	By their nature, VPNs are ill-suited to filter content. They function by establishing a secure connection to a private network.	VPN access is a choke point for connection to the network. Thus, as with ISPs, a VPN block would be of low cost to the provider.	And, again, as with ISPs, denial of VPN access would come with a high social cost of also denying non-malicious content. The negative cost is mitigated somewhat by the ease of diversion to another VPN system.	Less coercive than ISPs because of the ease of diversion to another VPN system. Unless VPNs maintained a global "ban" list, evasion of a VPN block would be comparatively easy.

	Knowledge of Risks	Better Risk Avoidance	Efficiency/Low Cost	Least Negative Social Cost and Bluntness of Tool	Most Coercive/Capable of Inducing Modification
DNS Operators	Low visibility into underlying content. Possibility that domain name may reflect malicious intent. Likely that malicious domain names are well known.	Poor choice for avoiding risk and/or sorting good content from bad. DNS registration is an on/off switch.	DNS operation is highly distributed. Changes in registration take time to propagate throughout the system. Once implemented, however, they are widely effective.	Social cost of completely ousting a user from the network is high. Registration under another domain name is, however, frequently feasible, rendering sanctions avoidable. Note however that the more extreme sanction of eliminating a DNS authorizer choosing to completely delist a gTLD or ccTLD (as was suggested early in the Ukraine-Russia war) would have even more severe negative social impact. In this circumstance, the intervention is not at all granular or precise.	De-registration of domain will have varying affect. Presumably gTLDs would not be impacted, so diversion to new domain name, while difficult, would not be impossible. However, settled costs of investment in an existing domain will create an incentive to modify behavior.

	Knowledge of Risks	Better Risk Avoidance	Efficiency/Low Cost	Least Negative Social Cost and Bluntness of Tool	Most Coercive/Capable of Inducing Modification
CDNs	Localized cached content could be reviewed if appropriate. Not currently subject to examination. May require legal change.	Cached content speeds up website resolution. Lack of access to a CDN does not, however, eliminate the underlying content, it merely slows down possible access. Thus, the risk of content is mitigated somewhat but not fully avoided.	Moderately high cost associated with tracking cached malicious content across a distributed network. Difficulty of individuated assessment. May be assisted by piggy-backing on routine CDN pushes to update caches. Necessity of appeals process increases cost.	In the absence of an appeals process, cost of false negatives would be high. Accuracy and efficacy of appeals process likely similar to that of currently operating moderation systems on social media platforms.	Multiple CDN providers are available. Lack of caching is effective at slowing access but not eliminating it. Coercive impact of sanction is thus limited.
Cloud Providers	Cloud service providers do not have regular visibility into content hosted on their servers. They do, however, monitor traffic for security purposes and could use that capability to review content, if legally authorized to do so.	Like ISPs, cloud providers could (if legally permitted) provide a reasonably effective venue for identifying malicious content.	Relatively lost cost of implementation. Good efficiency. Will require a costly appeals system, however. But in as much as the volume is less than for individual social media posts, costs of appeal would, likewise, be lessened.	Reasonably high negative impact from the possibility of false-negatives. A corollary of the effectiveness of coercion is the heightened possibility of adverse impact.	Access to cloud services is a near-essential part of the current information eco-system. Sanctions by cloud providers would be effective.

	Knowledge of Risks	Better Risk Avoidance	Efficiency/Low Cost	Least Negative Social Cost and Bluntness of Tool	Most Coercive/Capable of Inducing Modification
Web Hosts	Web hosts could have closer, retail-level knowledge of customer content and could more readily take action if they were aware of malicious activity.	Will have to create and provide a mechanism for moderating content. Likely to have smaller volume to review.	Distributed nature makes it likely that content moderation at scale will be costly and inefficient. Will also require an appeal mechanism.	Modest adverse social impact. Diverse nature of this part of the ecosystem will allow for alternate means of communication and for alternate venues for outlet of non-malicious content.	For the same reason sanctions will have only a modest coercive impact and will be easily evaded.
DDOS Mitigation	No direct knowledge of underlying content. But inferential knowledge is feasible from knowledge of threat models.	Virtually no organic capacity to mitigate risk. On/off switch of service provider is a blunt instrument.	Highly efficient at low cost for those impacted by the decision to remove protection. When your DDOS protection is removed the impact is strong and immediate.	The impact is also comprehensive. The entire web site/server structure is impacted including both malicious and beneficial content.	Highly coercive. In contemporary environment absence of security mitigation is a near-death sentence for a well-trafficed web-site.

	Knowledge of Risks	Better Risk Avoidance	Efficiency/Low Cost	Least Negative Social Cost and Bluntness of Tool	Most Coercive/Capable of Inducing Modification
Web Browsers	Web browser systems resolve HTML language into content. Current systems have pervasive tracking capabilities on web page interactions. That capability could be repurposed to review content.	As with social media, web browser systems will be able to identify malicious content at a granular level. To the extent the rule set can be hard wired into browser resolution it may actually be easier to scale than for social media moderation, though it will always need human back up.	Operation at scale may be easier than for social media. Rule set development will be costly and a human review/system will be necessary.	Significant possibility of adverse social impact through over-moderation. Faulty moderation decisions will need a robust appeal process.	Coercive capacity is high, especially if the moderation concept is adopted by all major web browser systems (otherwise avoidance will be possible). We may anticipate the possible development of unmoderated web browser systems whose use would mitigate the coercive impact.

	Knowledge of Risks	Better Risk Avoidance	Efficiency/Low Cost	Least Negative Social Cost and Bluntness of Tool	Most Coercive/Capable of Inducing Modification
Social Media Platforms	High degree of knowledge of risks with direct access to and visibility of malicious content.	Capable of identifying malicious content using artificial intelligence and human systems. Difficult to scale.	Very high costs of operation at scale. Requires significant human intervention and appeals process.	If the appeals process is effective, negative social costs through over-moderation are capable of mitigation. Where appeals process fails, significant negative social costs imposed. In either event, moderating a single comment is relatively precise and granular.	Comparatively easy to avoid. Yet significant impact in preventing large-scale communications (see, e.g. Trump ban from Twitter) has likely deterrent effect.
Online Marketplaces	Online marketplaces such as Amazon or Alibaba directly curate and algorithmically rank content. They could readily make themselves aware of malicious content if they chose to do so – and indeed many already do, excluding, for example, obscene material.	Though good as far as they go, online marketplaces are not the primary venue for malicious content. Thus content moderation here will have a less comprehensive impact than in other, more pervasive parts of the ecosystem.	Product review is individuated and difficult to scale appropriately. Can be assisted by automated sorting with human review but still comparatively costly.	Only moderate adverse social impact. Product providers will have other outlets for their speech – restrictions will simply make that speech more difficult to disseminate.	Significantly coercive. In ability to sell a product will defeat the very purpose of offering it. Limitations in the market store will incentivize producers developers to comply with marketplace content requirements (a circumstance we already see to some extent).

	Knowledge of Risks	Better Risk Avoidance	Efficiency/Low Cost	Least Negative Social Cost and Bluntness of Tool	Most Coercive/Capable of Inducing Modification
Search Engines/AI Curated Responses	Direct knowledge of search terms and search results. Except to the extent that harmful content is purposefully obscured by the user, there is significant visibility into queries by users and the responsive content being provided.	Excellent ability to identify risks. Indeed, interpretation of content is the hallmark of search engine capability. To be sure (as the idea of prompt engineering makes clear) monitoring content of search queries is not a complete solution – but content moderation through search engines would identify much malicious content.	Efficiency and cost are relatively small. Most significant limitation is that not all malicious content is accessed through search engines making coverage incomplete. We have also seen that limitations can be spoofed, allowing possible evasion of system.	Substantial possibility that search result editing may over-correct and edit out non-malicious content. But as with access to malicious content, it seems likely that the impact can be mitigated through re-configuration of searches and the intervention can occur with substantial granularity.	Search engine optimization is a business for a reason. Being dropped from a search engine and/or demoted down to a latter page has significant impact and will likely incentivize compliant activity.

	Knowledge of Risks	Better Risk Avoidance	Efficiency/Low Cost	Least Negative Social Cost and Bluntness of Tool	Most Coercive/Capable of Inducing Modification
App Stores	App stores (e.g. Google Play and Apple) curate apps made available based on anticipated content. They are also moderately well positioned to receive and respond to complaints of app non-compliance with content limtiations.	App stores have limited direct coverage of access to malicious content. To the extent apps are a useful gateway tool to the network, requiring them to have content limitations will impact access – but direct access will still be feasible.	App review is a relatively high cost endeavor. Often app functionality is not readily discernable and substantial analytical investment is required. Other aspects of an app may be more evident, but overall, this is a challenge.	As with online marketplace access, app store access is not a direct restriction on permissible speech. App developers will, generally, have other outlets for their content.	Significantly coercive. If nobody can download your app, nobody can use it. Will incentivize app developers to comply with app store content requirements (a circumstance we, again, already see to some extent).
Financial Services	Very low knowledge of risks from the underlying content. Engagement after the information has been exchanged/transaction has occurred.	Removal of financial incentives is a very indirect method of addressing malicious content.	Efficient method of addressing problems but with difficulties of scaling to meet individual case needs.	High impact – defunding a system does not prohibit speech *per se* but it will significantly impact the availability of content that can no longer be paid for.	Highly coercive. Though not all malicious content is intended to be offered for economic compensation, much is. To the extent the profit motive is eliminated, changed behavior is significant.

It is, perhaps, worth pulling out from this chart a few examples to explore in greater depth, as a way of illustrating the analysis more concretely.

Consider, first, web hosting services like Blue Host. These services would be attractive candidates for moderating malicious content as they typically have direct contact with, and therefore knowledge of, their customer's activity and content. They are, therefore, relatively well placed to act if they are aware of malicious activity, in much the same way that social media platforms are.

The first challenge, however, would be one of implementation. Currently, web hosting platforms do not undertake this level of scrutiny. Consequently, the hosting service community would have to create a mechanism for moderating content out of whole cloth. The cost of such an effort would be significant, especially since the distributed nature of web hosting services (there are more such services than there are, say, sizable social media platforms) would likely make content moderation at scale comparatively inefficient.

On the other hand, this diversity would mitigate any adverse social impact. Given that the web hosting system is more diverse than the social media platform system there would be several methods of communication and alternate venues for outlet of non-malicious content if a false positive were experienced. For the same reason, however, sanctions in the web-hosting system would have only a modest coercive impact and would be more easily evaded. And so, on balance, web hosting services would be an attractive, but imperfect venue for content moderation.

By contrast, to take another example, it seems clear that CDNs would be a relatively unattractive option. To be sure, localized cached content could be reviewed if appropriate, but that sort of examination is not currently very common, and doing so would likely require changes in law to authorize such activity.

More importantly, content moderation at the CDN level is likely to be ineffective. Caching content is useful to accelerate website resolution. But lack of access to a CDN does not eliminate the underlying content, it merely slows down a customer's access to the information. Thus, CDN moderation may mitigate the risks from malicious content, but those risks are not fully avoided.

In addition, it seems likely that there would be a relatively high cost associated with tracking cached malicious content across a distributed network. Unlike with web browsers, there would be significant difficulty in making individuated assessments of cached content. And, as with web browsers, the diversity of CDN providers would make sanctions ineffective and the coercive impact would be limited.

And so, as can be readily seen, least-cost analysis does not offer an easy answer to the content moderation question.

This challenge is further exacerbated by three other factors. The first is the absence of quantitative data. While the qualitative assessment we have made is, we believe, accurate, it is by no means certain. In the absence of financial data, the least cost principle can only serve as a guide to thought. It cannot reasonably be looked to for regulatory guidance.

Second, as should be evident, the least-cost avoider principle is an idealized examination of a theoretical construct. In the real world (even if accurate quantitative data were available) the scope of political expediency would be uncertain. To put it bluntly, were anyone to think seriously about imposing moderation obligations on search engines (another area where our analysis suggests fruitful inquiry) we

could expect that Google and Microsoft (the operators of Google Search and Bing) would oppose the effort. What is "right" from an economic perspective may not be "possible" politically.

Third, even leaving aside data availability and political economy questions, the complexity of the ecosystem, by itself, creates a significant barrier to assessment. There are many companies, such as Cloudflare and AWS, that provide services across multiple dimensions within the ecosystem. In the end, though the analysis is focused on service-specific inquiries, the reality is that obligations and responsibility will be applied at the platform- or enterprise-specific level. The disconnect may be inherent in the nature of the ecosystem and impossible to resolve.

Conclusion

Nevertheless, the analysis conducted makes two things clear -- not only are there a multitude of actors in the information ecosystem, but there are a number of them who, at least plausibly, may be thought of as well-suited to moderating content in ways that are akin to those currently employed by social media platforms.

Second, the analysis also makes clear that our current system of principal reliance on social media platforms may well prove ill-advised. Online marketplaces and app stores have already begun to take steps to moderate content. Other venues such as search engines and web hosting systems are also plausible venues for mitigating the harm from malicious content.

As policymakers move to regulate content moderation and/or mandate actions, they would do well to keep in mind the diversity of the information ecosystem and the possibility of broader and equally impactful interventions elsewhere in the system. In short, our least-cost avoider analysis does not give us a clear-cut answer, but it does suggest that the answer we have settled on – to rely primarily on social media platforms for content moderation – is both overly simplistic and, in the end, counterproductive.