ASAD RAMZANALI

Why the Cloud Needs Competition

To ensure the nation's leadership in artificial intelligence, cloud infrastructure must be regulated to encourage competition and enable innovation, building on lessons from other sectors.

Cloud computing is inextricably tied to the metaphor of clouds, those fleeting gatherings of water vapor that shift across the sky, capturing imaginations as they change shape and catch the light. Clouds are whimsical, ethereal, and notoriously impossible to pin down. Associating cloud computing with clouds invokes the sense of an invisible, ubiquitously productive presence floating over the earthly realm.

But cloud computing is actually grounded in hard, physical infrastructure. Cloud computing happens in data centers, which are giant warehouses of humming servers, each filled with semiconductors linked by hundreds of miles of copper wires and fiber-optic cables, cooled by millions of gallons of water per day, and powered by hundreds of megawatts of electricity. This real-world infrastructure enables nearly every digital activity, including opening an app, sending an email, or reading this essay online. Even the print edition of this magazine depends on the cloud: The US Postal Service runs hundreds of its operational tasks using cloud services.

Cloud computing's association with the whimsy of clouds obscures its vast physical infrastructure. It also obscures the industry's market concentration and profit model. Just three major providers control two-thirds of cloud computing revenues: Amazon (29%), Microsoft (20%), and Google (13%). This level of market concentration recalls industries like banking, telecommunications, and air travel, which are heavily regulated.

Still, policymakers have so far failed to regulate the cloud industry. This is a mistake that must be corrected. By providing an essential infrastructure on which the economy and national security depend, cloud computing has become an essential utility. But a lack of market competition may be stifling innovation. To ensure that the nation can lead in artificial intelligence and other emerging technologies central to US geopolitical and commercial interests, the cloud must be regulated in concrete ways, building upon lessons from the past.

Into the mega-blob era

Most people don't think of Amazon, Microsoft, and Google as cloud companies. Typically associated with e-commerce, business productivity, and web search, these vast conglomerates have many consumer-facing presences, including a grocery store chain (Whole Foods), a medical provider (One Medical), gaming platforms (Xbox, Activision Blizzard), social networks (LinkedIn), web browsers (Edge, Chrome) search engines

(Google.com), and many more. But for each company, the cloud business is foundational to its overall financial model. Amazon Web Services (AWS) accounts for 66% of Amazon's profits. Google Cloud's 34% growth in annual revenue is more than double its parent company Alphabet's overall growth rate of 16%. And Microsoft Cloud represented 63% of the company's revenues in its Q1 earnings release for fiscal year 2026. Earnings releases every quarter going back to 2020 recognize Microsoft's commercial cloud services as a major driver of revenue.

As venture capitalists, companies, governments, and other investors pour trillions of dollars into AI companies, many observers have declared we are in the middle of an AI gold rush. And as the old chestnut goes: To get rich during a gold rush, sell shovels and pickaxes. In the AI gold rush, the chip and cloud companies are raking in the money. OpenAI, for example, expects an annualized revenue rate of $20 billion (with no profits in sight)—considerably less than the revenues of just the last two months for the highly profitable cloud businesses of Microsoft and Amazon or the chip businesses of Nvidia and TSMC.

By providing an essential infrastructure on which the economy and national security depend, cloud computing has become an essential utility. But a lack of market competition may be stifling innovation.

Where this boom differs from previous Silicon Valley booms and busts is the interconnectedness of AI companies with their infrastructure providers. The economic exuberance of the AI boom is often compared to the dot-com bubble, when investor hype over the internet's potential funneled money into unproven startups like Pets.com and Webvan, which drove further investment from telecommunications companies as they rushed to expand networks to meet expectations for future web traffic. In that era, although early internet companies and infrastructure providers were both betting on industry growth, they occupied distinctly separate sectors.

Today's AI boom, by contrast, involves companies entangled, interlinked, and integrated vertically across the layers of the AI technology stack—chips, clouds, models, and applications. The largest cloud operators are often called hyperscalers because their business is in providing computing infrastructure at an enormous scale to meet demand. But the three leading hyperscalers' ambitions reach far beyond the cloud. Google designs its own AI accelerator chips (tensor processing units); owns Gemini, a popular AI model that Google uses in its own products and that third-party AI developers buy access to; and it owns the market-dominant search engine where AI is used—along with other key applications that use AI, such as Google Maps, Google Ads, and Google Drive. Google also makes its own smartphone (Pixel), mobile operating system (Android), and browser (Chrome) used to access AI applications, along with an internet service provider (Google Fiber) and electricity company (Google Energy), which provide key inputs for cloud computing data centers. Likewise, Microsoft designs some of its own chips, has the second largest cloud, operates its own AI models, and controls business applications where AI is being used (Office, Outlook, GitHub). Microsoft also has an energy company. Amazon designs chips, has the largest cloud operation, offers its own AI models (like Nova Act), and controls key applications where AI is used (e-commerce, smart home devices).

The tech stack has taken the concept of vertical integration—in which a company owns part or all of its supply chain—to new heights, allowing already massive companies to become largely impenetrable to conventional approaches to antitrust regulation. AI reflects this concentration and ties it directly to global financial markets and, by extension, the financial security of billions of individuals. Nine of the 10 most valuable companies in the world are focused on AI. And these companies are now critical to the US economy: In the first half of 2025, investment in AI data centers and computing infrastructure supported 92% of the growth in US gross domestic product, according to some estimates.

The AI boom is fueling further entanglements and interlinkages between hyperscalers and AI companies. Microsoft is one of the largest investors in OpenAI, and all three major hyperscalers are significant investors in Anthropic, a major competitor of OpenAI in the AI model market. Nvidia, the largest chip designer and the company with the highest market capitalization, is dabbling in its own cloud services and invested in upstart cloud provider CoreWeave. Nvidia is also becoming one of the largest investors in OpenAI (which itself is in the process of buying up part of AMD, an Nvidia competitor). If you're confused about how all this fits together, you're not alone. *Bloomberg* has called this a "web of circular deals," and Axios described this moment as "AI's mega-blob era."

Critical dependence

In late October 2025, both AWS and Azure, Microsoft's cloud computing platform, experienced minor outages caused by human error. The AWS outage took down "wide swaths of the web," *WIRED* reported, including WhatsApp, Signal, Reddit, and Snapchat. Some consumers couldn't access their bank accounts during the outage; others couldn't check their cameras to see who was at the door or change the temperature and position settings of their smart beds. The Azure outage stopped gamers from using Xbox and Minecraft, and it also took down the websites of Capital One, Alaska Airlines, Kroger, Starbucks, and Costco. Ironically, the Azure outage even degraded some of Microsoft's webpages that report its products outages.

These recent incidents demonstrate how cloud outages can cause societal disruption—but they are far from the only examples. Previous cloud outages and vulnerabilities have negatively affected hospital operations, delayed flights, enabled bank account breaches, and accelerated Russian hacking operations into US government websites.

A Congressional Research Service report described cloud computing as "analogous to the way electricity, water, and other utilities are provided to most customers." The cloud is like a utility because computing is an essential, commodified service, generated at a remote location (i.e., a data center) and delivered through a network (i.e., the internet). Like many other utilities, the industry tends toward concentration due to certain market features: high barriers to entry, economies of scale, and high switching costs.

Investigations by regulators in the United States and several Asian and European countries have built up an evidence base to prove how certain features of the cloud computing market (high market concentration, vertical integration, vast conglomeration) have allowed hyperscalers to engage in harmful corporate practices: opaque pricing of services for new customers, startups, researchers, and government agencies, for example, or artificially high cost of switching providers by charging egress fees when customers transfer data out of a cloud. These practices weaken markets, competition, and resilience.

Despite this recognition, it has proven difficult to gather the political will to regulate the cloud. Cloud infrastructure remains invisible to most voters—and so do its limitations and vulnerabilities—stymying the political movement that would be necessary to take on an industry worth three quarters of a trillion dollars.

How to regulate the cloud

In addition to being critical to nearly all facets of American life, cloud computing also undergirds frontier AI capabilities, which are key to modern geopolitical competition between the United States and China. But the goals behind winning the AI race are not about having bigger data centers or more capable chips—they're about controlling a technology that can enable leaps in productivity and scientific advancement.

Innovation in the applications of AI are dependent on infrastructure. If hyperscalers can pick winners and losers in AI applications based on their parochial interests, some uses of the technology may be squashed before they take off. Contrary to claims that regulation will stifle innovation, regulation is absolutely necessary to prevent market stagnation. A hands-off approach to regulating the cloud risks locking the country into a rigid and uncompetitive computing infrastructure with potentially disastrous consequences in the future.

Federal lawmakers can mitigate the security risks and economic harms of the cloud market by using a number of legislative levers. These actions constitute a package of policies that are separable but work best together as a regulatory regime—any single policy is useful and can work on its own, but in concert, this package lays the foundation for meaningful economic and national security regulation of cloud computing.

The most efficient approach would be to require Amazon, Microsoft, and Google to divest from AWS, Azure, and Google Cloud, respectively. Legislating such a structural separation has precedent; the tactic was used to separate railroads from owning coal mines, banks from commerce, and telecommunications and energy firms from adjacent markets. Applying structural separation to the cloud would prohibit hyperscalers from operating both the underlying computing infrastructure—especially the infrastructure- and platform-as-a-service (IaaS and PaaS) products—and the AI models, applications, and software-as-a-service (SaaS) products that depend on them.

Such divestments could actually be financially beneficial. Some investors have argued that spinning off cloud companies could increase shareholder value by removing the "conglomerate discount" and offering customers a "pure play" cloud investment opportunity (i.e., investors can currently invest in the three conglomerates but not the cloud market).

Another tool lawmakers can leverage is to institute neutrality rules for the cloud that could penalize hyperscalers for using price, access, or service discrimination to harm competitors. In general, neutrality rules combine two principles: an obligation to serve all customers on reasonable terms, and a nondiscrimination requirement to treat them equally. Nondiscrimination rules have been used in the United States to govern railroads, telegraphs and telephone lines—though in recent years courts struck down an effort to apply those telecommunications neutrality laws to broadband providers. For the cloud, neutrality rules should be

designed to prohibit anticompetitive practices that are common today, like committed-spend discounts, egress fees, bundling unrelated products, copying customers' tools, or prioritizing affiliated applications. Neutrality regulations often also require transparency around pricing, service quality, and denial-of-service policies.

A third opportunity is to require data portability and interoperability, which would lower switching costs and reduce lock-in. The Telecommunications Act of 1996 did both, requiring phone number portability across carriers and mandating that providers' networks enable interconnection. Portability allows customers to freely move data from one provider to another, while interoperability requires different systems to exchange data interactively. Applied to the cloud, legislation should require providers to enable timely, no-cost data transfers and to adopt open-source standards for real-time interoperability. Together, these policies would make multi-cloud operations feasible and create a more open, resilient market.

Fourth, boundaries can and should be legislated around foreign ownership of US cloud computing companies. Foreign ownership of US companies has long been restricted in industries that are central to national security—banks, broadcasters, airlines, nuclear plants, and now even TikTok. Cloud computing should be no exception. Legislation could establish clear thresholds, such as prohibiting 25% of ownership, profits, or board control by foreign companies, and require that directors of major cloud service providers be US citizens. These measures would align cloud policy with long-standing safeguards for other critical industries; they would also be responsive to increasing trends in other countries, including US allies, toward calling for "sovereign clouds" to avoid dependence on American companies.

A fifth action is to require cloud service providers to follow know-your-customer (KYC) obligations to verify customer identities to prevent misuse. Recently, policymakers have proposed the concept to semiconductor exports, but cloud providers face similar risks: Anonymous access to large-scale computing can enable malicious cyber operations or unauthorized AI training. The Commerce Department's 2024 proposal to apply KYC rules to IaaS providers builds on executive orders from presidents Trump and Biden. As in the financial services industry, KYC rules would help protect a critical system from exploitation without impeding legitimate use.

Finally, policymakers can designate cloud computing as critical infrastructure at the federal level. Sixteen sectors, including energy, emergency services, chemicals, and financial services, are designated critical infrastructure by the Department of Homeland Security. The designation signifies these sectors as being "so vital to the United States that the incapacity or destruction of such systems and assets would have a debilitating impact on security, national economic security, national public health or safety, or any combination thereof." Given its systemic role in every other sector, from hospitals to defense, cloud computing warrants a standalone critical infrastructure designation. Doing so would enable greater threat-intelligence sharing, asset tracking of data centers, and coordinated responses to outages and cyber incidents.

Today's AI boom involves companies entangled, interlinked, and integrated vertically across the layers of the AI technology stack—chips, clouds, models, and applications.

2026 marks the twentieth anniversary of Amazon's launch of AWS, which kicked off modern cloud computing. Two decades later, the term *the cloud* still shields this crucial infrastructure—and the vastly increased complexity of the industrial and financial engineering buttressing it—from deeper scrutiny and regulation. Over time, the cloud market appears to be ossifying in ways that could hinder competition today and innovation tomorrow. And, because it is central to American markets and geopolitics, that rigidity could further threaten US economic and national security. Global competition around AI only intensifies these hazards. Looking forward, decisionmakers should ensure that market failures do not circumscribe future innovation to the narrow path determined by today's cloud providers. Only foolhardy prospectors would try to mine for gold using blunt pickaxes and shovels. Market governance through tried-and-tested regulation for stimulating competition is a crucial step to sharpen this country's tools for the AI gold rush.

Asad Ramzanali *is the director of artificial intelligence and technology policy at the Vanderbilt Policy Accelerator at Vanderbilt University.*