Developing and Deploying AI Responsibly:
Elements of an Effective Legislative Framework to Regulate AI

Written Testimony of Brad Smith
Vice Chair and President, Microsoft Corporation

U.S. Senate Judiciary Committee
Subcommittee on Privacy, Technology, and the Law

September 12, 2023

Chairman Blumenthal, Ranking Member Hawley, Members of the Subcommittee, thank you for the opportunity to testify on what I believe is a critical issue for the Congress and the country. I am Brad Smith, the Vice Chair and President of Microsoft Corporation.

I welcome the opportunity today to share some initial thoughts on the proposed Blumenthal-Hawley framework and the critical role legislation must play in ensuring effective oversight of artificial intelligence (AI).

In short, I believe this framework is a strong and positive step towards effectively regulating AI. It reflects the urgency and speed needed to address this fast-moving technology, and it combines strong protection for the public with support for ongoing technology innovation.

Importantly, the framework builds on other federal efforts, like the White House AI commitments unveiled in July and the bipartisan AI Insight Forums, providing the constructive interplay needed between the executive and legislative branches. And it will enable Congress to listen and learn before addressing every detail, while establishing not just a blueprint but the initial foundation on which additional legislative steps can be taken.

As the legislative process moves forward, I hope Congress will include three goals in the list of priorities that deserve the most attention.

First, Congress should prioritize AI safety and security. As I discuss in more detail below, the Blumenthal-Hawley framework addresses these needs in a strong manner, including by proposing a licensing regime under an independent oversight body with a risk-based approach for AI models and uses. Microsoft supports this approach and believes it strikes a sensible balance that can both protect the public and advance innovation, even while we recognize the need to work through a large variety of hugely important details. By also incorporating transparency and security requirements, the Blumenthal-Hawley framework puts Congress on a path to provide the public with the safety standards it deserves.

Second, Congress should ensure that AI is used in a manner that complies with longstanding legal protections for consumers and citizens. This should include the protection of privacy, civil rights, and the needs of children, as well as safeguards against dangerous deepfakes and election interference. The Blumenthal-Hawley framework rightly addresses these issues head-on, without shying away from the critical issues that Congress must consider if it decides, as we believe it should, to replace Section 230 with new approaches that are a better fit for AI technology.

The framework also focuses, as we believe it should, on the differing roles of AI developers and AI deployers, in effect creating a pragmatic regulatory architecture that reflects relevant AI technology architecture. The Blumenthal-Hawley framework makes room for considering in a practical way where current laws may be sufficient if enforced well, as well as for the possibility of new rules where needed. In sum, it strikes a sensible balance based on a practical understanding of the relevant technology, the need to advance innovation, and the imperative to protect the nation's citizens.

Third, Congress should ensure that AI is put to good use to build a government that can better serve our citizens. We should not overlook the opportunity to put AI to use as a valuable tool, even while we protect against its potential abuse as a technological weapon. Some national governments in other countries are moving quickly to use AI to improve healthcare, strengthen education, make public services more accessible, and advance public sector efficiencies. In the United States, state leaders like Governor Newsom in California, Governor Burgum in North Dakota, and Governor Youngkin in Virginia are taking an early lead in using AI to build better state governments. At the federal level, we should consider the role of legislation and oversight to advance similar goals, as well as the strengthening of the country's national defense and security. I hope the Blumenthal-Hawley framework will expand to make more room to address not only new risks associated with AI, but new opportunities as well, especially to build a government that can better serve the public.

In the sections that follow, I focus in greater detail on the principles that Microsoft believes should guide the development of legislation to govern AI, many of which are embraced by the Blumenthal-Hawley framework. This includes ensuring that legislation promotes accountability for both AI development and deployment, builds on existing work, and reflects the technical architecture of AI itself. I will also discuss priority areas where federal regulation and oversight seem particularly appropriate, including requiring "safety brakes" for highly capable AI models used in critical infrastructure, and requiring developers of AI systems to know their customer, their cloud, and their content.[1]

1. Promote accountability in AI development and deployment

When we at Microsoft adopted our six ethical principles for AI in 2018, we identified one principle that should serve as the bedrock for all the others: accountability. Accountability means that AI systems must be subject to effective oversight by humans. It also means that the people who develop and deploy these systems must remain accountable to everyone else, affording those impacted by harmful AI systems protection under the rule of law.

We therefore welcome the decision by Senators Blumenthal and Hawley to make accountability a centerpiece of their proposed regulatory framework. It recognizes that to govern AI effectively, we need to act in a targeted way and place the right expectations on the right stakeholders to address the risks of greatest concern. Developers and deployers of AI systems must work together to strengthen AI safety and apply special care in the highest risk scenarios; AI systems that are used to make consequential decisions about the most vulnerable members of our communities should be subject to greater oversight than the AI systems that help us find the next emerging musician based on our prior playlist.

As this Committee considers how best to ensure that the United States remains a world leader in responsible innovation, we would encourage you to focus not only on accountability in the development of AI, but also on accountability in its deployment. Promoting accountability in deployment means risk will be managed more effectively, and that AI can be put to work more broadly to help people in their day-to-day lives and make progress on our greatest societal challenges.

Throughout history, we have seen that countries that lead in the uptake of new technologies are often the ones that fare best, sometimes even over countries that may be more technologically innovative. Consider, for example, the printing press. Few inventions in history have had such profound effects on the world. Yet although the printing press was invented in Germany in the

[1] Further details on these ideas can be found in Microsoft, *Governing AI: A Blueprint for the Future* (May 25, 2023), at https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RW14Gtw.

1400s, it was the Dutch and the English who first truly embraced printing and books. Indeed, by 1500, citizens in the Netherlands were reading more books per capita than anyone else in the world. It is certainly no coincidence that England and the Netherlands also soon found themselves at the forefront of economic innovation and global commerce.

This lesson is worth bearing in mind as we think about how governments can best promote the benefits and responsible use of AI. Although AI innovation is undoubtedly important, those countries that succeed in rapidly *adopting and using* AI responsibly are the ones most likely to reap the greatest benefits. Fortunately, the United States is well placed to lead both responsibly and rapidly because of its existing AI safety initiatives.

2. Build on existing efforts

The United States is the home of many of the world's top developers of advanced AI. Given the foresight of U.S. policymakers and regulators, it also has a number of existing AI safety frameworks on which to build—including the White House initiative to secure voluntary commitments from industry, and the NIST AI Risk Management Framework. These initiatives target different aspects of AI development and deployment, making them especially relevant to use in concert to support accountability across both activities. As this Committee considers the proposed Blumenthal-Hawley framework and other efforts already underway in Congress to regulate AI, we encourage you to take account of these existing initiatives.

The White House voluntary commitments focus on three fundamental principles: safety, security, and trust.[2] They target developers of 'frontier' AI models – the most advanced models that exceed the capabilities of currently released systems like OpenAI's GPT-4 – requiring participating companies to take on specific duties designed to advance each of these principles, in particular:

- Ensure that their products are *safe* before offering them to the public. This includes commitments to engage in red teaming of frontier models to identify safety risks, to share information with other companies and governments on emerging risks and threats, and to develop standards and best practices for frontier AI safety.

- Build systems that put *security* first. This includes commitments to invest in cybersecurity and insider threat safeguards, and to provide incentives for third parties to discover and report AI security issues and vulnerabilities.

- Do right by the public and earn people's *trust*. This includes commitments to deploy provenance technologies or watermarks so that people know when they encounter AI-generated audio or visual content; to publicly disclose model or system capabilities, limitations, and appropriate and inappropriate uses, including effects on fairness and bias; to prioritize research into societal risks posed by AI systems; and to develop and deploy frontier AI systems to help address society's greatest challenges.

Microsoft was among the first companies to adopt these commitments, and I am proud to say that we have gone beyond them[3]—a point I will come back to later in this testimony. Among other

[2] *See* White House, *Voluntary AI Commitments*, at https://www.whitehouse.gov/wp-content/uploads/2023/07/Ensuring-Safe-Secure-and-Trustworthy-AI.pdf.

[3] *See* Microsoft, *Voluntary Commitments by Microsoft to Advance Responsible AI Innovation* (July 21, 2023), at https://blogs.microsoft.com/wp-content/uploads/prod/sites/5/2023/07/Microsoft-Voluntary-Commitments-July-21-2023.pdf.

things, we have committed to implement the NIST AI Risk Management Framework (AI RMF) across our own AI development and deployment practices, and to attest to this with our customers.

NIST developed the AI RMF based on the directive that this Congress issued in the National Artificial Intelligence Initiative Act of 2020. The framework, which supports risk management efforts by both developers and deployers of AI technologies, builds on NIST's years of experience in developing similar frameworks to address cybersecurity risks. We commend NIST for having developed the AI RMF, and in particular for having done so in a transparent and consensus-driven process involving input from both the public and private sectors.

We are pleased to see that the Blumenthal-Hawley regulatory framework incorporates transparency measures to promote responsibility and due diligence from the companies developing and deploying AI systems, and we believe the NIST AI RMF could be a useful tool to inform the risk management strategies and practices about which they should be transparent. One place to start could be federal procurement. The Federal Government has a proven track record of using procurement rules to incentivize industry to innovate, improve their products, and embrace industry best practices more generally.

Building on the model of existing rules that require federal contractors to adopt strong cybersecurity practices, Congress could likewise encourage industry adoption of standards based on the AI RMF by requiring federal contractors to self-attest, as a condition of bidding on federal contracts, that they have implemented those standards. Because it's always better to learn to walk before trying to run, Congress could initially focus on the procurement rules for critical decision systems, meaning AI systems that have the potential to meaningfully impact the public's rights, opportunities, or access to critical resources or services. As industry uptake of responsible AI practices increases, the Government could expand the scope of these procurement rules to additional areas as appropriate.

Congress could also consider directing NIST to establish an AI RMF program office to provide federal agencies with guidance on the framework and to promote its adoption. This Office could also provide resources on AI trustworthiness to officials responsible for procuring such systems. In tandem, the General Services Administration and OMB could be directed to develop voluntary, standard contract language for agencies to use in those procurements.

While the White House voluntary commitments and the NIST AI RMF provide a strong foundation, we also need laws and regulation that build on this foundation and complete the AI regulatory architecture.

3. Require safety brakes for AI that controls or manages critical infrastructure

Artificial intelligence is not the first-time societies have confronted a valuable new technology that also has the potential for harm if it fails. History is replete with examples.

For instance, the growth of cities led to increasingly taller buildings, which in turn required the use of elevators. People were understandably worried about what might happen if the cables holding an elevator aloft were to fail. Elisha Otis, the inventor of the elevator, solved this by developing a brake that would catch the elevator car before it fell. It wasn't long before city building codes required in the installation of safety brakes on all elevators. And it worked: today, elevators are supremely safe, and we don't give a second thought to stepping into one.

That pattern has repeated itself many times. Governments require circuit breakers in buildings to protect against fires caused by surges in electricity. They require school buses to have emergency

brakes, in case the main brakes fail, and require bus drivers to be trained in how to use them. They require airplanes to have collision avoidance systems installed, and to ensure that pilots are able to make decisions based on those systems in safety-critical situations.

The common thread running through all these examples is that, where a technology failure can cause significant or widespread harm, it is often appropriate to require suppliers to install back-up safety systems, and to ensure that people have the ability to use them if they're ever needed.

The Blumenthal-Hawley framework recognizes this critical role for 'safety brakes'. A clear scenario in which they are needed is when highly capable AI models are used to access, manage, or control critical systems, the failure of which could cause widespread harm—for example, systems controlling the power grid, first responders, transportation traffic, and the like. Laws requiring developers to build safety brakes into such systems, and requiring that deployers can use them effectively, would promote accountability by ensuring that these systems remain under human control at all times.

Although the details of what such a regulatory regime would look like deserve further consideration, we envision it having at least four components:

- First, direct regulators to define the class of high-risk AI systems controlling critical infrastructure that would require safety brakes. At least initially, regulators could focus on highly capable AI systems that: (i) take decisions or actions affecting large-scale networks; (ii) process or direct physical inputs and outputs; (iii) operate at least semi-autonomously; and (iv) would pose a significant potential risk of large-scale harm if they were to fail.

- Second, require AI developers to build safety brakes into the design of these designated AI systems. Although the type of 'safety brake' would likely vary depending on the system and how it was used, all of them should have the ability to detect and avoid unintended consequences, and to disengage or deactivate the AI system in the event of unintended behavior.

- Third, require deployers of designated systems to test and monitor them to ensure that they remain within human control. Although exact requirements will depend on the system, deployers should have to periodically test, verify, and rigorously validate the operation and performance of the system and its components, consistent with safety best practice. Deployers should be accountable for proving that they can operate safety brakes built in by developers and that the system remains under human control at all times.

- Fourth, AI systems that control the operation of designated critical infrastructure should be deployed only in licensed AI infrastructure. Because a point of vulnerability for AI systems can be the computing infrastructure on which they run, we think it is also worth requiring the operators of such infrastructure to be licensed. To obtain a license, the operator would need to design and operate their infrastructure in a manner that allows another point of intervention—in effect, another safety brake in the event that the application-level measures fail.

Today, there are relatively few AI systems that exert the kind of control over critical systems that would necessitate these types of safety brakes. But we shouldn't use that as an excuse for delay. Given the rapid pace at which highly capable AI systems are progressing, we should begin developing the rules of the road now for the future that we know is coming.

4. KY3C: Know your customer, cloud, and content

As the Blumenthal-Hawley framework recognizes, it is important to think about how laws can help promote the responsible use of AI *and* guard against its misuse. It's also valuable to leverage regulatory frameworks that have proven themselves to be effective in addressing risks that share characteristics in common with those we are most concerned about for AI.

In the financial services context, 'know your customer' obligations seek to minimize the risk that financial institutions unwittingly facilitate transactions that are being used for illegal ends. They have been critical to advancing U.S. national and economic security, and the integrity of the U.S. financial system, by enabling authorities to identify and tackle money laundering, terrorist financing, and other crimes.

More recently, a refined set of 'know your customer'-inspired techniques is emerging as a best practice in the cybersecurity space. Leading cloud service providers maintain policies and processes to protect against threat actors gaining inappropriate access or engaging in certain abusive activities, including by using digital indicators and machine learning to hunt for threat actors, disable their accounts, and assess new accounts for fraud and abuse risk.

In our view, a version of this framework could be applied with good effect in the AI context, advancing several of the goals of the Blumenthal-Hawley framework. In particular, we think it's worth considering a "KY3C" regulatory framework that would impose obligations on various actors in the AI value chain, requiring them to know their *customers*, their *cloud*, and their *content*. Specifically:

- Know your customer. At least in scenarios involving high-risk or other sensitive use cases, it might be appropriate to require operators of the cloud infrastructure on which a highly capable AI model is running to know the customers who are accessing the model. Depending on the scenario, it might also be appropriate to require AI model developers or operators of highly sensitive AI applications to know their customers, managing access not only to AI datacenter infrastructure but also to the models and broader capabilities for sensitive uses. As in the financial services sector, the goal would be to reduce the risk that AI providers unwittingly let bad actors use powerful AI models or applications running on their infrastructure to engage in theft, fraud, or other misuse.

- Know your cloud. Developers of highly capable AI models, in turn, should have an obligation to "know the cloud" on which their models are deployed. This obligation would help reduce the risk that bad actors can exploit vulnerabilities in that cloud to corrupt the functioning of those models. To satisfy this obligation, developers would have to use licensed AI cloud infrastructure, requiring the cloud provider to meet ongoing regulatory requirements proving that they have taken the necessary steps to protect that infrastructure against malicious attacks and adversarial actors.

- Know your content. Although the creative potential of highly capable AI models is astounding, it also opens the door for the creation of deepfakes, misinformation, and other malicious content. To help reduce the risk of people being deceived or misled, we think the public deserves to "know the content"—that is, to know when they come into contact with images or audiovisual content that has been produced or altered by AI. The law could require deployers of AI systems to accomplish this through the use of provenance technologies or watermarking, at least in scenarios presenting a significant risk of harm.

> Key technical building blocks to enact such a legislative requirement already exist. One of the most important of these is the Coalition for Content Provenance and Authenticity (C2PA), co-founded by companies including Adobe, the BBC, Intel, Microsoft, Sony, and Truepic. As Microsoft's Chief Scientific Officer Eric Horvitz has said, the success of such labeling will require education aimed at media literacy, awareness, and vigilance, together with investments in quality journalism. Although success on these fronts will require work, the health of democracy and meaningful civic discourse will undoubtedly benefit from initiatives such as these that help protect the public against deception or fraud facilitated by AI-generated content.

We appreciate that these proposals will require further thought and discussion, including on how best to balance national security and crime-prevention goals against the need to protect privacy, freedom of expression, and other fundamental human rights. But if we target regulation at those use cases that pose the greatest risk, and put adequate safeguards in place, we think it is possible to balance these interests in ways that benefit both individuals and society.

5. Ensure the regulatory framework maps to the technology architecture of AI

Both the safety brake and the KY3C proposals discussed above are examples of a broader point—namely, that the regulation of AI needs to map to the technology architecture of AI itself. To be effective, the law needs to place different regulatory responsibilities on different actors based on their role in managing different aspects, or layers, of AI technology. This is particularly important when it comes to regulating highly capable and other frontier AI models, which is a critical aspect of any federal regulatory regime. It's also consistent with the Blumenthal-Hawley framework's embrace of the core principle that both developers and deployers should be accountable for AI safety.

There is no single right way to describe this technology architecture, and different engineers might well describe it differently. For our purposes, it's useful to think of AI as involving a technology stack made up of three core layers: the application layer, the model layer, and the infrastructure layer, more commonly known as the "cloud" layer. Each layer of this technology stack, in our view, necessitates a distinct regulatory approach. AI developer accountability at the infrastructure, model, and application layers must be implemented in ways that correspond with the technology stack, just as AI deployer accountability at the application layer must also correspond with the risk management capabilities uniquely available as AI is deployed.

The applications layer

At the top of the technology stack are AI-powered applications and services. These are what deliver information and other AI outputs directly to users. The application developer may operate the AI "model"—the prediction or classification engines that power these applications and services—itself, or obtain that functionality from a third party through an "application programming interface," or API. OpenAI's ChatGPT, Microsoft's Bing Chat, and GitHub's CoPilot (an application that generates software code in response to user prompts) are all examples of AI-powered applications. Microsoft also offers APIs to enterprise customers that allow them to build applications on top of AI models that Microsoft makes available from OpenAI and Microsoft itself. This lets enterprises easily build AI solutions to suit their own needs, without having to develop their own AI models or build the sophisticated infrastructure on which these models run.

Because this is the layer at which people directly interact with AI outputs, it is also the layer at which the safety and rights of people will be most impacted. As a result, we need to ensure that the laws and regulations that currently govern conduct apply with equal force to those who provide or deploy

AI applications and services. Simply put, we need to apply and enforce the laws that are already on the books.

For example, the fair performance of AI systems across different demographic groups is a well-known and pressing concern for AI development. It is already unlawful for a bank to discriminate against a mortgage applicant based on race or gender. If a bank decides to use AI tomorrow to help it evaluate loan applications, it will need to ensure that this doesn't lead to such discrimination. We don't need new laws to make this happen; we just need to apply the laws we already have.

Although this may sound simple, the implications are potentially profound. Those who develop and deploy AI applications and services will need to think hard, and proceed carefully, to ensure that they fully comply with existing laws—most of which, however, were drafted before the advent of AI. Regulators will need to know how to apply those laws to AI applications and services, and courts will need to know how to interpret and enforce those laws with regard to AI.

The model layer

The next layer down in the AI technology stack consists of pre-trained AI models. These are the foundational technologies that power the top layer of AI applications and services. Most AI models used today are designed to do discrete tasks—such as translate text, or recognize patterns in images—and do not, in our view, require new legislation. However, a class of highly capable AI models is emerging that may require new regulatory approaches, both for the models themselves and for the infrastructure on which they run.

These highly capable AI models are unique in many respects. They are typically trained on huge, internet-scale datasets and include billions of parameters that interact in exceedingly complex ways. Many are also effective out-of-the-box at doing a wide range of tasks—from drafting a mathematical proof or writing a poem, to developing software code for a new application or creating an image of a sunset seen from Mars.

In addition to being multifaceted and powerful, the behavior of these frontier, highly capable AI models can also be hard to predict. In many cases, these models are so complex that the outer bounds of their capabilities can only be determined in practice, such as through controlled releases with users. As a result, harnessing the full potential of these models, while also ensuring that they align with our laws and values, is challenging, and our understanding of how to do this effectively is evolving.

Although developers like Microsoft are addressing these risks through rigorous testing and self-imposed standards, the risks involved are too important, and their scale and potential impacts at present too unknowable, to address them through self-regulation alone. We therefore think it is appropriate for Congress to consider legislation that would impose a licensing regime onto developers of this discrete class of highly capable, frontier AI models, and we are pleased to see that the Blumenthal-Hawley regulatory framework seeks to establish such a regime.

Although the details of this licensing regime again would benefit from further thought and discussion, and there are critical consequences and details to deeply consider, such as the impact to open source models and the importance of continuing to foster an innovative open source ecosystem, we think it should seek to serve three key goals:

- First and foremost, any licensing regime must ensure that the development and deployment of highly capable AI models achieve defined safety and security objectives. In concrete terms, this

may require licensees of these models, among other things, to engage in the pre-deployment testing that the Blumenthal-Hawley regulatory framework proposes. We agree that highly capable models may need to undertake extensive prerelease testing by internal and external experts. In addition, a licensing regime may require developers of highly capable models to provide advance notification of large training runs; engage in comprehensive risk assessments focused on identifying dangerous or breakthrough capabilities; and implement multiple other checkpoints along the way.

- Second, it must establish a framework for close coordination and information sharing between licensees and regulators, to ensure that developments material to the achievement of safety and security objectives are shared and acted on in a timely fashion. The Blumenthal-Hawley framework provides that an independent oversight body not only conducts audits but also monitors technological developments, which may be best accomplished in partnership with licensees. The adoption of controls over model deployments, potentially based on the assessed level of risk and evaluations of how well-placed users, regulators, and other stakeholders are to manage residual risks, may be required. Post-release monitoring may also help ensure that the models are functioning as intended and remain under human control at all times.

- Third, it must provide a footing for international cooperation between countries with shared safety and security goals. Because AI systems and their outputs are not confined by geographic borders, domestic regulation alone will not be enough to secure the benefits of highly capable AI models and guard against their misuse. We believe there is an opportunity for the United States to work with like-minded countries to advance an international framework for AI governance, enabling an AI system evaluated as safe in one jurisdiction to qualify as safe in another. There are many effective precedents for this, such as common safety and security standards set by the International Civil Aviation Organization, which allows for an airplane to fly from Brussels to New York without being re-fitted over the Atlantic. The United States could also work with others to advance a global consensus on the most pressing risks and opportunities around AI and to improve our collective understanding of AI safety.

The cloud infrastructure layer

The third layer in the technology stack is the cloud infrastructure layer on which highly capable AI models are developed and run. This infrastructure is more powerful than the datacenters that run more traditional digital applications and services: it typically provides far greater computer power, uses specialized AI chips, and involves significant engineering skills and resources.

Because it provides the technological foundation on which highly capable AI models run, this AI cloud infrastructure layer is a key control point for these models—and also a potential point of vulnerability if they are not managed properly.

We therefore see a role for licensing providers of this cloud infrastructure to ensure that they act responsibly in ensuring the safe and secure development and deployment of highly capable AI models. To obtain a license, an AI datacenter operator would need to satisfy specified technical capabilities around cybersecurity, physical security, and safety architecture. As noted above, it would also include a “know your customer” requirement to guard against these datacenters unknowingly permitting their infrastructure to run AI applications used for criminal or other harmful purposes, and have the capability to provide a “safety brake” on AI systems used to control or manage critical systems.

These requirements and a new generation of export controls can help protect U.S. national security interests and avoid the proliferation of frontier models to adversaries and those intending to cause harm. The need for such regulation will become increasingly clear as AI models on the frontiers of technology become more capable, more autonomous, and more likely to bridge the digital-physical divide. Congress would do well to ensure that we stay ahead of these risks by ensuring the appropriate legislative guardrails and authorities are in place ahead of time.

6. Conclusion

Powerful new AI technologies should give all of us grounds for optimism, given their many potential benefits when they are developed and deployed responsibly. At the same time, we must not ignore their potential perils. Industry plays an essential role in promoting the safe and responsible development of AI. But laws and regulations have a vital role to play as well. At their core, these laws should require AI systems to remain subject to human control at all times, and ensure that those who develop and deploy them are subject to the rule of law. We need to place the right expectations on the right stakeholders to address the risks of greatest concern. The Blumenthal-Hawley framework sets consideration of all of this on the right course.