Submitted March 15, 2024

CCIA Comments on Singapore's Proposed Model Governance Framework for Generative AI

Introduction & Summary

Below please find the submission of the Computer & Communications Industry Association ("CCIA") regarding Singapore's Infocomm Media Development Authority (IMDA) and AI Verify Foundation's proposed "Model AI Governance Framework for Generative AI: Fostering a Trusted Ecosystem" ("Proposed AI Governance Framework").[1] CCIA is an international, not-for-profit trade association representing a broad cross section of communications and technology firms. For more than 50 years, CCIA has promoted open markets, open systems, and open networks.[2]

CCIA appreciates Singapore's consultative approach to developing these guidelines and how the Proposed AI Governance Framework generally avoids imposing rigid constraints on AI developers and deployers. Singapore's proposal generally aligns with best practices, which rely more on the flexible and international technical standards-based approaches to governance of AI that are crucial to supporting AI innovation and diffusion.[3] This is particularly important as AI develops quickly and due to the general-purpose nature of AI technology. Imposing overly-prescriptive rules while the technology still develops could slow innovation and would likely become outdated quickly as global standards are themselves still in development and constantly changing along with AI technologies and their applications in the real world. To further these recommendations, attached to this submission is a June 2023 report prepared by CCIA detailing recommendations for non-disruptive and effective AI governance titled, "Understanding AI: A Guide To Sensible Governance."[4]

Overall, CCIA appreciates Singapore's broadly-supportive and non-restrictive approach to AI governance, particularly as many jurisdictions have pursued inflexible definitions and broad-sweeping categorizations of risk that could hinder the development and deployment of AI technologies in those markets.

However, the Proposed AI Governance Framework can use further improvement to ensure it is maximally clear, flexible, and effective for actors in the AI ecosystem. CCIA provides specific comments on Sections 1, 2, 4, 5, and 7 of the Proposed AI Governance Framework below.

Accountability

First, the Proposed AI Governance Framework correctly focuses on accountability as a key aspect of AI governance. However, CCIA suggests that accountability should be applied throughout the AI lifecycle from development to deployment, with the specific accountability

[1] https://aiverifyfoundation.sg/downloads/Proposed_MGF_Gen_AI_2024.pdf.
[2] For more, visit www.ccianet.org.
[3] https://ccianet.org/library/understanding-ai-guide-to-sensible-governance/.
[4] https://ccianet.org/library/understanding-ai-guide-to-sensible-governance/.

mechanisms assigned to each stage of the lifecycle tailored based on proportionality. Specifically, end-users who misuse or abuse AI technologies should be held accountable themselves—such oversight should not be reserved for developers, who may have little control over end users maliciously using what may otherwise be a well-designed AI system. CCIA therefore suggests that the Proposed AI Governance Framework be amended to include end-user accountability.

Second, assigning responsibility to developers for mitigations and corrections would be a more constructive way to implement accountability mechanisms than through user rights for redress, particularly for a burgeoning ecosystem of technologies. Given the different parties involved– developers (which, for open-source models, could involve multiple parties), deployers, and end users–assigning liability for the effects of a model can be complex. The current Proposed AI Governance Framework draft stipulates that "there should be consideration for how responsibility is allocated both upfront in the development process (ex-ante) as best practice, and guidance on how redress can be obtained if issues are discovered thereafter (ex-post)."[5]

The context of Generative AI and its vast potential to improve individuals' lives, catalyze economic growth, and strengthen governance should be incorporated into Singapore's AI governance. Given that Generative AI is a nascent technology, oversight and governance of this form of AI should focus on mitigation of anticipated harms resulting from the applications of this technology rather than introduce overarching "redress" mechanisms for scenarios that have not yet been analyzed by technologists, policymakers, civil society, and businesses. Such flexibility would give regulators and individuals experiencing harmful consequences of Generative AI the ability to fix those problems as they arise, while not introducing a vague and potentially burdensome perpetually looming threat of users calling for redress.

Third, the definition adopted in the Proposed AI Governance Framework for "open source" should be expanded to ensure a broad set of current and future AI systems are not excluded inadvertently due to an overly-rigid set of restrictions. Currently, there is no global consensus on a definition of "open source," and many different interpretations of the term exist in the technology industry. As such, governments seeking to oversee AI technologies have generally sought to avoid cementing any one interpretation of "open source" to ensure they do not freeze their regulations or guidelines at one moment in time and potentially benefit one interpretation of what "open source" is over others.

The specific definition selected in this Proposed AI Governance Framework is problematic, as it suggests AI developers using an "open source" model must allow for end users to access the "full source code and information required for re-training the model from scratch."[6] While in some cases, an AI developer may choose to use a fully open source model, others will employ hybrid approaches in which only some of the source code and/or model is accessible to end users. Further, open source model developers may train their model on data they are permitted to use but not to share, making it impossible to comply with this requirement. In addition, AI developers may need to restrict access to various portions of their system to

[5] https://aiverifyfoundation.sg/downloads/Proposed_MGF_Gen_AI_2024.pdf at 6.
[6] https://aiverifyfoundation.sg/downloads/Proposed_MGF_Gen_AI_2024.pdf at note 7.

ensure responsible use—i.e., to prevent end users from leveraging "open source" models of AI to engage in illegal or malicious activities. Given that the Proposed AI Governance Framework as currently drafted also does not assign accountability to end users using licensed models, this narrow definition could further undermine reasonable practices of AI developers.

CCIA recommends that the Proposed AI Governance Framework acknowledge open source systems as a spectrum that includes varying degrees of openness. As such, CCIA would point to the definition put forward through the European Union's 2020 Open Source Software Strategy as a more conducive approach to labeling such models: "Open source software (or free software) combines copyright and a licence to grant users the freedom to run the software, to study and modify it, and share the code and modifications with others. It facilitates collaboration, innovation, and agility."[7] This definition would not only broaden the scope of Singapore's interpretation of "open source," it would also ensure that the Proposed AI Governance Framework would not need to distinguish between "open source" and "open weights" models.

Fourth, deployer responsibility should not be directly connected to developer reputability, as such an approach could ignite geopolitical industrial and security debates and could lead to a bureaucratic registration obligation.[8] CCIA recommends against linking the responsibility and accountability of AI players to AI developers' reputability, as such a gauge is arbitrary and vague and as such would be difficult to implement. Determining what constitutes "reputable platforms," as put forward in the Proposed AI Governance Framework, differs between countries and is an inherently subjective exercise, particularly absent clear international consensus and related means to demonstrate conformance, when necessary. This is especially true in the modern geopolitical context: economic and security rivalries and fights are often based on technological innovation and regulation, and some jurisdictions could label competing models from foreign markets as harmful or disreputable. If the government of Singapore is obligated to discern which platforms are reputable, it could put the IMDA in an uncomfortable position which jeopardizes its neutrality between numerous third-party countries. Further, the development of criteria that establishes reputability for this responsibility could result in what would effectively be a certification requirement for AI models that would discourage AI deployment.

Fifth, given the lack of distinct delineations between the different actors in the AI ecosystem in Singapore's approach, CCIA believes it is premature to introduce indemnities and insurance in the Proposed AI Governance Framework. Despite the importance of incorporating shared responsibility in AI governance, it is crucial to exercise careful consideration before leveraging economic concepts such as indemnities and insurance at this early stage of emerging technologies such as AI systems. This is particularly concerning given the following circumstances, all of which are fundamental to the nascent nature of the AI ecosystem and make the implementation of terms such as indemnities and insurance premature:

---

[7] https://commission.europa.eu/document/download/97e59978-42c0-4b4a-9406-8f1a86837530_en?filename=en_ec_open_source_strategy_2020-2023.pdf.

[8] https://aiverifyfoundation.sg/downloads/Proposed_MGF_Gen_AI_2024.pdf at 6 ("Responsibility in this case, for example when using open-source/weights models, should require application deployers to download models from reputable platforms to minimise the risk of tampered models.").

- The players that are central to the AI ecosystem have not yet fallen into definitive roles and responsibilities;
- The risks and how they should be allocated to these players in the AI ecosystem have not been studied extensively; and
- There is insufficient study into the informational imbalances and the possible externalities that may be present in the AI ecosystem.

Finally, CCIA recommends the removal of the reference to the European Union's AI Liability Directive in the Proposed AI Governance Framework as it has not yet been finalized and takes a problematic approach to governance.[9] The proposed revision to the EU's Product Liability Directive and proposed AI Liability Directive have not yet been enacted. Criticisms of these proposals include potentially disincentivizing European AI development and the wider innovation ecosystem in the bloc. As such, CCIA recommends against using the proposed AI Liability Directive and Revised Product Liability Directive as an example for Singapore's regulatory plans, as it is too early to know how these proposals would impact innovation and consumer welfare in the European Union and the potential application to Singapore's market.

Instead, Singapore could use the method pursued by the White House's Voluntary AI Commitments[10] to foster AI actors to demonstrate accountability in the following ways:

- "Work toward information sharing among companies and governments regarding trust and safety risks, dangerous or emergent capabilities, and attempts to circumvent safeguards";
- "Invest in cybersecurity and insider threat safeguards to protect proprietary and unreleased model weights";
- "Incent third-party discovery and reporting of issues and vulnerabilities"; and
- "Develop and deploy frontier AI systems to help address society's greatest challenges".

Data

CCIA recommends the removal of the clause about seeking consent from copyright owners for use of their material for Generative AI development in the Proposed AI Governance Framework,[11] as Singapore's copyright law precludes such an obligation. Singapore's Copyright Act, revised in 2021,[12] does not require AI developers to seek consent to use copyrighted material for training data, and a further amendment to the current copyright laws would be an unnecessary intervention that could have detrimental effects on the broader digital sector. Singapore's Copyright Act includes an exception for text and data mining

[9] https://ccianet.org/wp-content/uploads/2023/05/CCIA-Position-Paper-on-AI-Liability-Directive.pdf.
[10] https://www.whitehouse.gov/wp-content/uploads/2023/09/Voluntary-AI-Commitments-September-2023.pdf.
[11] https://aiverifyfoundation.sg/downloads/Proposed_MGF_Gen_AI_2024.pdf at 8 ("From a model development perspective, the use of copyright material in training datasets and the issue of consent from copyright owners is starting to raise concerns. Models are also increasingly being used for generating creative output – some of which mimic the styles of existing creators and give rise to considerations of whether this would constitute fair use.").
[12] https://sso.agc.gov.sg/Act/CA2021.

(referred to as "computational data analysis") along with a "fair use" exception. In combination, these two provisions suggest that the proposed mention of protecting copyright holders in the context of data accessibility need not be included. This approach would not preclude rightsholders from availing themselves of an exclusion protocol, such as Google's Google-Extended protocol,[13] and thereby opting out of text and data mining of their content, should they so choose.

Incident Reporting

As Singapore considers approaches to incident reporting, CCIA recommends against holding up the European Union's AI Act as an example, given it was only recently finalized and its effects have not yet materialized. As such, the reference to the AI Act should be removed from the Proposed AI Governance Framework. The impacts of the AI Act will likely not be known for years—the law grants industry a transition period of around two years to prepare to comply. Once compliance is in effect and the law is enforced, it would likely take at least a year—and potentially up to two years—before stakeholders coalesce on an understanding of whether the threshold for incident reporting is appropriate and whether the structures and framework of the reporting are effective.

Further, without having yet been implemented, CCIA urges caution regarding the AI Act's proposed 15-day period for reporting serious incidents, as it is logistically not feasible to execute a meaningful investigation of an incident in that time. Such a rigid and quick timeline for incident reporting would greatly compromise the quality of incident reporting.

Testing and Assurance

CCIA supports the inclusion of testing and assurance duties in the Proposed AI Governance Framework, as responsible AI developers generally already center such obligations as part of their own internal practices. However, the Proposed AI Governance Framework should include a broad definition of "testing," instead of restricting it to "red-teaming" to encapsulate a wide range of testing methods, such as fit-for-purpose measures. Red-teaming can require large amounts of resources, including significant amounts of worker hours and money. These costs associated with red-teaming testing—particularly if applied to all AI models, irrespective of the potential risks that could be posed by the model—are prohibitively high and could discourage AI innovation.

The Proposed AI Governance Framework references the White House Voluntary AI Commitments noting that "several AI companies pledged to conduct external model red teaming as a means of demonstrating trust."[14] CCIA supports linking Singapore's governance to the these industry commitments, as the agreed-to red-teaming obligations are limited to frontier models, which are defined as "models that are overall more powerful than any currently released models, including GPT-4, Claude 2, PaLM 2, Titan and, in the case of image generation, DALL-E 2."[15] Setting guardrails such as those secured by these voluntary

[13] https://blog.google/technology/ai/an-update-on-web-publisher-controls/
[14] https://aiverifyfoundation.sg/downloads/Proposed_MGF_Gen_AI_2024.pdf at note 22.
[15] https://www.whitehouse.gov/wp-content/uploads/2023/07/Ensuring-Safe-Secure-and-Trustworthy-AI.pdf.

commitments for mandatory red-teaming testing would be critical to ensure the time, money, and resources necessary to execute such methods would be reserved for appropriate scenarios.

Further, CCIA recommends that IMDA consult with AI actors and other government entities to provide a set of specific risks, harms, and content for which AI developers should be testing. AI developers must have a robust understanding of the potential harms they are attempting to root out while testing, as open-ended queries for vaguely-defined negative outcomes are likely to be unsuccessful due to the effectively infinite possibilities that can be tested against. A strong understanding of the subject area is key to AI developers identifying what to search for and test against. Subject matter experts in arenas such as biology, chemistry, and other specialized subjects are difficult for most technology firms to assess, as they do not have the necessary expertise in-house. However, such knowledge likely exists in the government in the inter-agency process. As such, CCIA urges explicitly including a note that AI actors would not be held liable for identifying harms or problem areas absent clear guidance from IMDA and/or other government agencies.

Content Provenance

Ensuring the ability for consumers to identify AI-generated content is important and CCIA supports its inclusion in the Proposed AI Governance Framework, in line with the commitments secured between the White House and AI actors in 2023. Content provenance tools that are used to label AI-generated content, like watermarking, should be limited to audio, photo, and video content that could be reasonably confused with synthetic content.

Provenance and transparency tools necessary to label generative audio and visual content are under intensive technological development seeking to address unsettled obstacles. The final AI Governance Framework should center obligations in content provenance for content that can be used maliciously to deceive individuals that are unable to identify it as AI-generated, such as “deep fakes” and should limit any requirements for content provenance to what is technologically feasible. The threat of deceiving or confusing consumers would necessarily not apply to content that can clearly be determined as AI-generated. As such, CCIA recommends the language from the White House’s Voluntary AI Commitments: “Audiovisual content that is readily distinguishable from reality or that is designed to be readily recognizable as generated by a company’s AI system—such as the default voices of AI assistants—is outside the scope of this commitment.”[16]

Further, any requirements for content provenance should not apply to text. The technology to embed watermarks into text is still developing and in some ways is a more complicated scenario than audiovisual and visual content. If companies were to include unremovable watermarks into AI-generated text, it could confuse individuals more than it helps, as many people use AI-generated text as a starting point for writing their own outputs. For someone who asks AI to generate text based on a prompt to help as part of their own process, the inclusion of a watermark could suggest that the entirety of their product is AI-generated when

[16] https://www.whitehouse.gov/wp-content/uploads/2023/09/Voluntary-AI-Commitments-September-2023.pdf at 3.

in reality, only a small portion of it would be. It is also difficult to embed text watermarks in a way that is not trivial to remove, limiting the value of this approach for text.

Conclusion

CCIA appreciates the opportunity to offer feedback on Singapore's developing AI Governance Framework. As the attached report details, as Singapore continues its work in this area, flexibility is paramount to effective and innovation-friendly governance of AI. The public interest can be protected while ensuring AI technologies are not hindered in their development and deployment into society.

Computer & Communications Industry Association

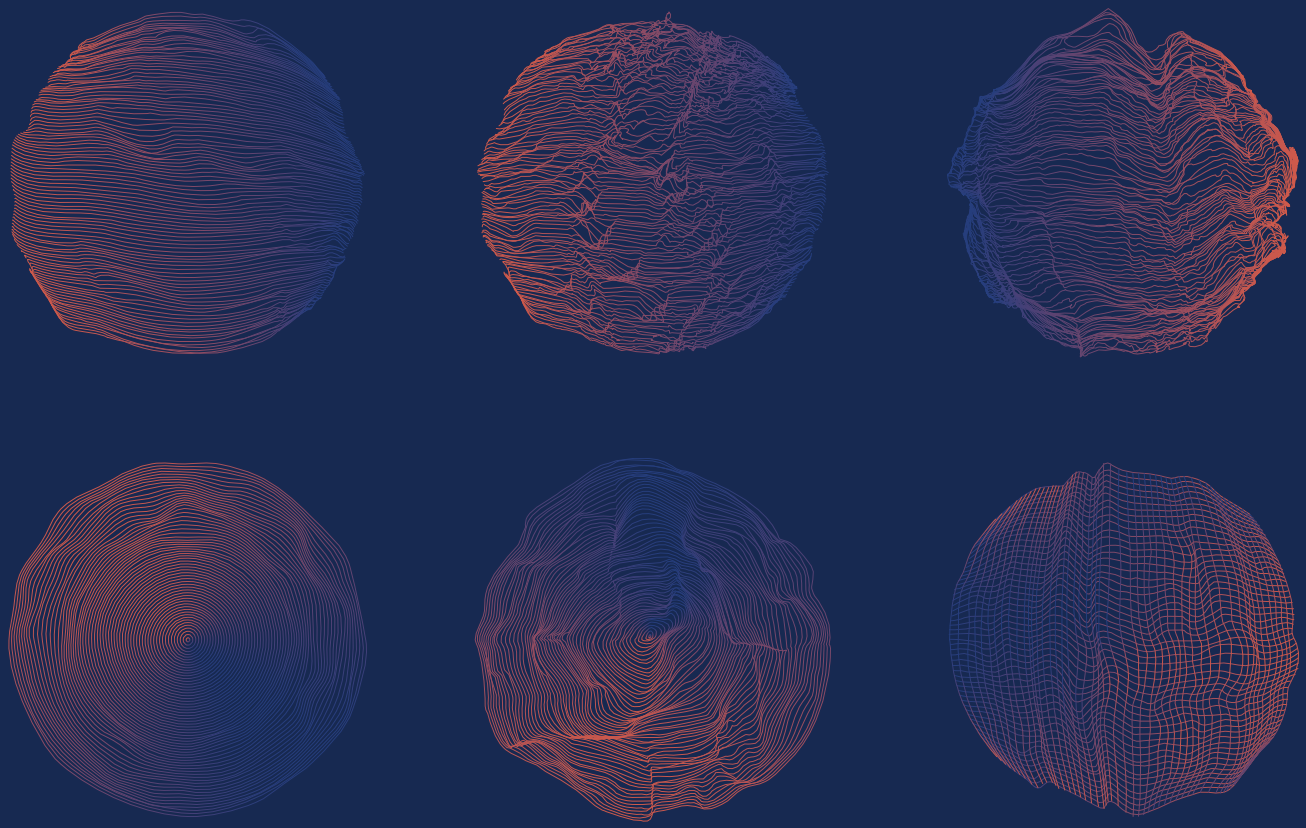
Open Markets. Open Systems. Open Networks.

ccianet.org

2023

Understanding AI

A Guide To Sensible Governance



Executive Summary

In today's rapidly evolving technological landscape, artificial intelligence (AI) has emerged as a powerful force with the potential to reshape various aspects of society, from economic prosperity to national security. However, only through careful consideration and a deliberate approach to regulation can we harness the benefits of AI and mitigate its potential risks. Critically, AI is not a single technology but rather a family of related, but distinct, technologies, each of which may be applied in significantly different contexts. Applying rules designed for one type of AI or one context to another situation can hinder the development of new forms of AI and create, rather than reduce, harms.

To ensure effective regulation and self-governance of AI, a multistakeholder approach is vital. Drawing from the successes of the broader internet governance ecosystem, a similar framework can be applied to AI governance. Such an approach allows for diverse perspectives, fosters innovation, and accommodates the evolving nature of AI technologies.

Existing laws can address aspects of AI that are not unique to the technology. Whether performed by a human or an AI, illegal discrimination already violates federal and state laws, for example. Allowing existing law to cover AI overall, while also identifying the limited instances where AI introduces unique challenges that may require discrete additions to existing law, will result in a predictable and stable environment for AI investment, limit duplicative regulation and regulatory arbitrage, and ensure that the benefits of AI flow to Americans while mitigating potential harms.

Regulation will also play a vital role in engendering trust in AI systems. By establishing clear guidelines and standards for transparency and accountability, regulation can help address concerns related to privacy, bias, and accountability. But overly prescriptive approaches, like those under the EU's AI Act, may hamper the development of the next generation of AI technologies. And regulation of AI can also create outcomes that are antithetical to the U.S. system of democratic institutions, as with China's draft law requiring AI services to obtain political pre-approval.

Regulating AI Requires Understanding AI

AI has already become an integral part of our lives. Technologies like speech and facial recognition and machine translation are forms of AI that are already widely used. While recently developed technologies like Large Language Models and transformer-based image generators have drawn recent attention, regulation of AI must avoid unintended consequences by taking into account these other forms of AI, as well as the rapid pace of advancement in AI technology. New types of AI are continuously being developed, making it challenging to predict the precise direction of advancement in AI technology. To foster innovation and progress, it is important not to implement rigid regulations that rely on the present mechanisms by which AI operates, but rather to take approaches that manage overall risk in a way that incorporates the context in which each AI system operates. One example of such an approach is the National Institute of Standards and Technology (NIST) AI Risk Management Framework, which was created per Congressional direction.

Among the existing types of AI, there are several prominent examples worth mentioning:

- **Automated Decision-Making (ADM):** Algorithms autonomously make decisions based on predefined rules and data. Existing practical applications of ADM are nearly endless, with ADM used in diverse fields from scaling content moderation tools to increasing access to financial credit.
- **Machine Perception:** Enables machines to understand sensory inputs. This includes computer vision and speech recognition. Practical applications of machine perception can be seen in Shopify's automatic product description generation, making it easier for businesses to create detailed product listings, and in accessibility tools that automatically describe images for visually impaired individuals.
- **Natural Language Processing (NLP):** A form of AI that focuses on machine understanding of human language. NLP is often combined with machine perception to enable a machine to interact with humans more naturally. Applications like Google Translate and natural language search engines such as Google and LexisNexis exemplify the capabilities of NLP, and voice assistants like Siri, Alexa, and Google Assistant apply a combination of NLP and machine perception to listen to, understand, and respond to human requests.
- **Machine Learning (ML):** A technique for creating various forms of AI, including some of those used in NLP or machine perception. ML involves training algorithms with large datasets to recognize patterns and make predictions or decisions. Generative models and Large Language Models (LLMs) are examples of ML-based AI systems that have gained significant attention recently. These models have demonstrated impressive capabilities in generating realistic text, images, and even entire stories.

While these applications of AI may not hold the same level of attention as recent generative AI tools, they have already solved real problems. Translation allows people to access documents that were created in languages they don't speak. Image recognition has been used to detect potholes in roads and to improve weather forecasting. And automated decision-making techniques have helped to modernize occupational license processing and to make water management decisions more quickly and with better outcomes. These existing applications hint at the tremendous potential AI holds, if implemented responsibly with appropriate risk management.

Developing AI Responsibly Requires Flexible Regulation

In the rapidly advancing landscape of AI, responsible development and deployment are paramount. However, it is crucial to strike a balance between regulation and flexibility, avoiding overly prescriptive principles that may stifle innovation. To achieve this delicate equilibrium, the principles of responsible AI should be considered in designing thoughtful, adaptable regulation that can be applied in all contexts. Rather than being overly prescriptive, the focus should be on designing AI systems for the benefit of society while proactively analyzing and mitigating risks during the development and deployment processes.

One significant consideration is guarding against overbreadth in definitions. Regulation should focus on high impact decisions where AI plays a crucial role. Clear delineations must be established to distinguish between AI as a contributing factor in decision-making and instances where AI makes decisions without human review. By doing so, we can ensure that appropriate oversight is in place while avoiding unnecessary constraints on AI development.

Similarly, caution should be exercised to prevent overbreadth in implementation strategies. Human guardrails may be beneficial in certain cases, providing necessary checks and balances. However, it is essential to recognize that no single approach will always be correct. Flexibility is key when determining the level of human involvement, ensuring that the level aligns with the unique characteristics and requirements of each AI system.

Broad agreement exists among leading AI developers and researchers, including CCIA's members, that responsible AI development requires the following:

- Design for social benefit.
- Design to avoid unfair outcomes.

- Analyze and minimize risks as you design.
- Consider the risks to third parties from AI systems during design, but also the benefits.
- Use up-to-date safety, security, and privacy best practices.
- Monitor and govern identified risks in deployed systems.
- Provide appropriate disclosures for deployed AI systems.

While these principles may be expressed in different ways, any responsible AI framework will incorporate them. CCIA's members have engaged in responsible AI development, ranging from developing and applying their own responsible AI principles to conducting academic research that promotes privacy-by-design and the hardening of AI against motivated attackers seeking to extract training data, among other valuable contributions.

These high-level principles, applied in the context of any given application, provide the necessary flexibility to manage risks while providing the benefits AI can deliver. In high-risk applications, such as medical diagnostics, human supervision and significant disclosure of the AI would be appropriate; in lower risk applications, such as content moderation or video games, there may be little or even no need for human review.

AI Warrants Only Targeted Regulation Combined With Considered Application Of Existing Law

Rather than rushing to create new laws, it is essential to evaluate whether existing laws at the federal, state, and local levels adequately address the concerns posed by AI. In general, there should be little to no difference whether an act is performed by a person or by an AI system. This can be achieved by writing and applying law and regulation in a way that constrains outcomes, while maintaining neutrality as to the process by which those outcomes are created. For example, instead of creating a new law requiring AI systems to operate in a non-discriminatory fashion, existing discrimination laws should be applied to AI systems. By leveraging established legal frameworks, we can address these types of concerns without burdening the regulatory landscape with unnecessary redundancy. Using established legal frameworks and applying them evenhandedly to AI and human systems alike will also avoid regulatory arbitrage by ensuring there will be neither a legal advantage nor a disadvantage to operating a system as an AI system versus via human action.

The effective application of existing laws, such as intellectual property (IP) laws and product liability laws, will also address the vast majority of concerns that have prompted calls for the regulation of AI systems. Recent statements by officials from the FTC, DOJ, EEOC, and CFPB emphasize exactly this approach. These technologically neutral laws should be the first line of defense, addressing common legal issues when they arise in the context of AI applications. But where AI-specific distinctions exist, or when a failure of existing law emerges, new regulations tailored to that unique situation should be created.

Moving Towards A Risk-Based Framework For AI

Comprehensive regulation of AI should employ a risk-based framework rather than a prescriptive framework requiring specific mechanisms. National standards such as the NIST AI Risk Management Framework and international standards such as ISO/IEC 23894 and ISO/IEC 42001 may be relevant to refer to in the development of risk-based approaches. Policy-makers should focus on identifying and addressing the concerns associated with AI development and deployment. This approach empowers developers to find appropriate solutions within the defined limits while not limiting room for new technologies and experimentation.

The level of acceptable risk, required guardrails, and potential impacts should be evaluated based on the specific context. For applications with lower impact, higher tolerable risk levels and fewer guardrails may be acceptable. Conversely, applications with higher impact demand lower tolerated risk and more robust guardrails. This approach allows flexibility and adaptability, catering to the diverse nature of AI technologies.

Appropriate levels of transparency and disclosure are also crucial aspects of AI regulation. While they may not impact benefits or harms, they are essential to engendering trust in AI systems. People should have access to relevant information about how an AI system was designed and trained, as well as how it operates. This knowledge fosters accountability and user trust, enabling individuals to understand the basis of AI-driven decisions.

While transparency is important, it must be appropriate and relevant. Context is the key factor in determining the needed level of transparency, with riskier AI systems requiring higher levels and potentially more human involvement. An AI system that directs the movement of pallets in a warehouse should require significantly lower levels of transparency than an AI system that makes lending decisions. Additionally, protection of proprietary knowledge and confidential

business information is critical. Striking a balance between transparency and confidentiality is vital to promote investment in innovation while maintaining ethical and accountable AI practices.

Addressing Specific Issues That Have Received Attention

A. Determining responsibility for AI outputs

There are a number of different entities involved in any given AI system, including the provider who trained the AI model, the deployer who applies that model to a specific task, the compute provider who provides the hardware the AI system runs on, and the user who ultimately is utilizing the AI system. Basic legal principles of agency can serve as a starting point for determining responsibility. The developer, deployer, user, and compute resources involved in an AI system might each bear responsibility, depending on the circumstances.

Compute resources, typically acting as intermediaries or common carriers, should generally not be held responsible for AI outputs. On the other hand, trainers of a model may be held accountable if defects are inherent to the design of the AI system. For instance, if a model developer intentionally creates an AI that consistently ranks people of color as less creditworthy, they should bear responsibility for that, not just the operator of the system.

Similarly, operators of AI systems, while they may be generally responsible for the usage of the technology, should not be held liable for inherent design flaws or the actions of users if users can interact with the operator. For example, if a user instructs an AI to generate defamatory content, the operator should not be liable for that content.

This division of responsibility will ensure that liability lies in the most appropriate place, with the actor most capable of minimizing harm and most responsible for any harms that ensue.

B. Determining regulatory responsibility

While a governmental coordination role might be useful, creation of a new department or similar bureaucracy is likely to lead to regulatory duplication and stifle investment in and development of AI systems. In most cases, existing agencies responsible for specific areas of law are equipped to oversee regulation of AI that falls within their area of responsibility. Leveraging the

expertise and jurisdiction of these agencies will ensure a coherent regulatory landscape. For example, housing discrimination law would fall under the purview of agencies like the Department of Housing and Urban Development (HUD) or the Office of Fair Housing and Equal Opportunity (FHEO).

Similarly, coordinating the regulatory efforts and fostering industry development of best practices across various domains could be the role of the National Institute of Standards and Technology (NIST) or a similar entity; another potential model is the role of the IP Enforcement Coordinator in the IP ecosystem. Such coordination ensures consistency of the overall approach while allowing domain experts to ensure effective regulation of AI systems within their agency's expertise.