Recommendations for Regulating AI

The rapid advance of AI technologies has spurred governments around the world to consider AI regulation. AI has the potential to deliver extraordinary productivity and economic gains, and make generational progress in science, healthcare, energy, and many other fields.

Governments that effectively harness AI's potential will gain a significant competitive advantage, positioning their countries to prosper and thrive.

A thoughtful AI opportunity agenda, as well as considered measures to address AI-specific risks, builds trust in technology and promote widespread adoption.

Getting AI regulation right is therefore a key public policy responsibility for every government. Decisions governments make today on the "**how**, **what**, and **when**" of regulation will profoundly influence the trajectory of AI innovation and adoption.

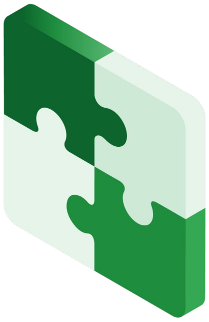
AI is too important not to regulate — and too important not to regulate well. Regulation must be grounded in evidence and understanding of the full potential of this rapidly evolving technology.

That's why Google supports well-designed, evidence-based regulation that fosters certainty and predictability within the AI ecosystem — for developers, for deployers, and for users. Smart regulation can promote responsible AI practices while ensuring that innovation can continue to flourish.

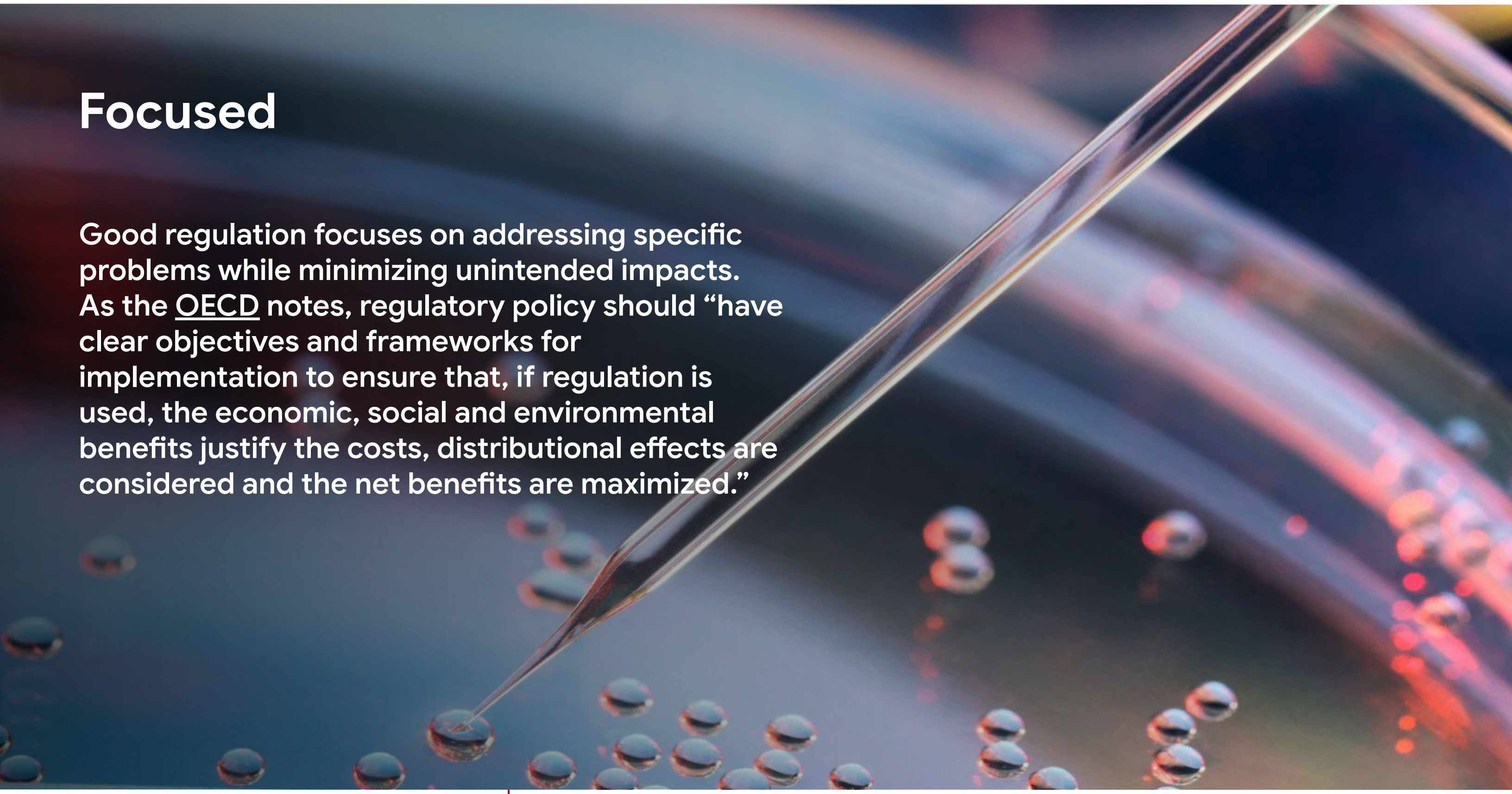
This paper consolidates our recommendations on designing good, pro-innovation AI regulation, which we believe should be **focused**, **aligned**, and **balanced**.

Focused

Aligned



Balanced



Focused

Good regulation focuses on addressing specific problems while minimizing unintended impacts. As the OECD notes, regulatory policy should "have clear objectives and frameworks for implementation to ensure that, if regulation is used, the economic, social and environmental benefits justify the costs, distributional effects are considered and the net benefits are maximized."

Regulate real-world effects, not scientific progress

AI is a general-purpose technology, like electricity or engines. It can be deployed in a wide variety of contexts, from translation applications to medical diagnostic tools.

The primary driver of risk is generally derived from the precise context of its use. For example, uses of AI in banking will differ from uses in pharmaceutical research or transportation. An AI assistant writing a grocery shopping list is lower risk than a system that helps a doctor diagnose pneumonia. Regulations at the level of foundational models cannot account for variations in risks and benefits that are driven by who is using AI, for what purpose, and in what settings.

It is also important to recognize that sectors generally recognized as *high-risk* are also often *high-value* – and there are uses in such sectors where AI applied responsibly can actually *reduce* risk. Medicine and transportation are sectors where AI has been shown to reduce human error rates and contribute to improved outcomes for patients and passengers.

Broadly regulating a general technology also tends to choke off beneficial, lower-risk uses of AI, while hamstringing innovation. It also misses the point of addressing risks and harms where they happen, and where end users and consumers interact with AI. Regulating use cases allows policymakers to fully leverage the benefits of AI while still protecting consumers, businesses, and society.

Identify and address regulatory gaps

In considering and designing regulatory interventions for AI use cases, the starting point should be a comprehensive review of existing laws that already apply. If it's illegal to do something without AI, it is likely already illegal to do it with AI.

For example, many countries have laws prohibiting discriminatory employment practices. These laws apply whether or not AI was used in the decision-making process. Similarly, existing laws concerning fraudulent practices and consumer protection apply whether or not AI is present. We should focus on whether we need to enforce existing laws more effectively; whether AI creates new issues not already addressed; and whether the use of AI creates a need for new obligations.

Legal gaps can arise in several ways:

(i) where there are no existing legal provisions – for example, where there is no law covering discriminatory hiring practices;

(ii) where laws exist but are inadequate to mitigate the potential concerns precipitated by AI – for example, where synthetic media may raise new issues for elections advertising; or

(iii) where laws exist but need to be changed in order to create an enabling environment for AI development, deployment, and adoption – for example, where we need to amend government procurement rules to facilitate AI adoption.

Once governments have identified the unique issues raised or significantly altered by AI technology, they can explore a range of regulatory options: issuing clarifications, guidelines, advisories, or amendments to existing laws.

New regulations risk reinventing the wheel, and imposing new costs for governments, for companies, and for consumers. Efforts should be proportionate and calibrated according to the degree of novel or enhanced risk, as well as real-world evidence of actual harm or other market failure.

Focus on the outputs

It is difficult for regulators to micromanage the rapid developments in computer science driving AI progress. Prescriptive requirements, particularly at the input level (such as the data sources for model pre-training), will quickly become outdated and fall behind the state of the art. Output-based requirements are more suitable for the governance of fast-moving technologies, focusing on real-world impacts while allowing industry to research and develop the best approaches to meet regulatory goals.

For example, privacy controls are more effective at the output and application level, where there may be not only greater potential for harm (such as greater risk of personal data disclosure), but also greater opportunity to use safeguards. Leakage of personal data, or hallucinations misrepresenting facts about a non-public living person, often happen through interaction with the AI application, not through the development and training of the base AI model itself.

At the same time, the output and application stage offers opportunities for privacy safeguards and protection against inappropriate, offensive, or harmful content. Options include enforcement of usage policies; limitations on how the application interacts with personal data; watermarks, filters, classifiers, and other output safeguards; and enhanced transparency and user controls.

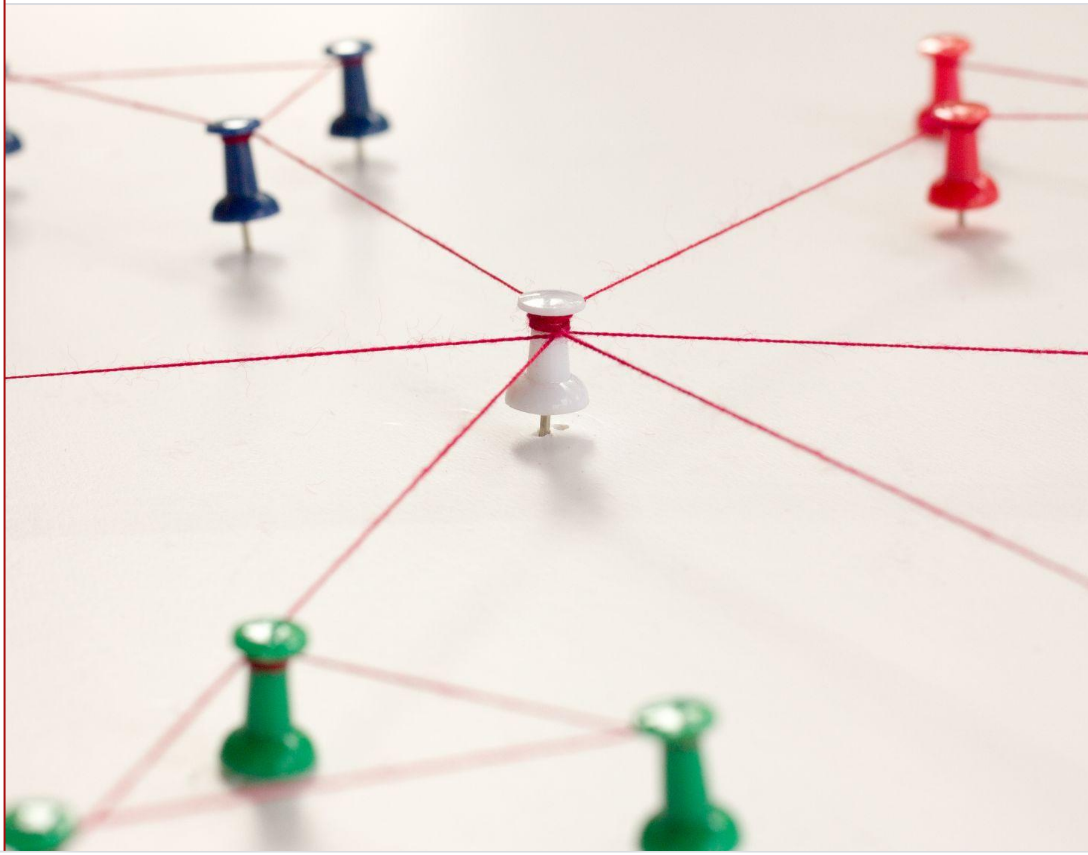
Consider a sectorally-based, hub-and-spoke governance architecture

We recommend a hub-and-spoke AI governance architecture, where a “hub” of centralized AI technical expertise is maintained to support sectoral regulators in their regulation of specific AI applications — the “spokes”.

The hub of centralized AI expertise can lay out sector-agnostic AI risk management best practices and frameworks, and provide consulting expertise within government for technical advice on the capabilities and implications of evolving AI technologies.

This function should complement the “spokes” — sectoral regulators tasked with regulating the use of AI in their specific areas of expertise. The risks of AI are highly context-dependent — issues in finance will differ from those in healthcare or transportation.

Sectoral regulators are best placed to assess context-specific uses and impacts of AI in their respective domains, and whether and how best to regulate them. For instance, health-focused agencies are best positioned to evaluate the use of AI in medical devices. Similarly, transportation agencies have expertise in evaluating the deployment of autonomous vehicles.



Encourage effective techniques to identify AI-generated content

Governments should consider targeted and effective requirements to promote trust, transparency, and the responsible development of machine-readable AI.

Provenance technology, like watermarking, fingerprinting, or embedded metadata, can help people better understand how a particular piece of content was created and modified over time. This helps people make more informed decisions about the content they're engaging with and builds media literacy and trust.

Of course, knowing whether a piece of content is AI-generated is not always the sole or even the most useful piece of information in empowering people to decide whether they want to trust it. Indeed, a piece of AI-generated content can be helpful, accurate, and informative–just as much as a piece of authentic content can be misleading or promote scams.

It is also the case that more and more content will be created using AI tools, just as most content today is created with computers. And as growing amounts of content are marked as "AI-generated", the marks may lose their significance as users develop "banner blindness," and stop noticing labels once they become ubiquitous.

Worse: overbroad use of visible labels may engender "implied truth" effects, whereby people would deem content not marked as AI-generated as more likely to be accurate or trustworthy.

But at least during this interim period, requirements for generative AI systems to include provenance data in image, video, and audio outputs can be a targeted and effective way of promoting digital literacy. In contrast, mandating requirements for prominent, user-facing labels on all AI-generated content could have unintended consequences, and actually undermine the goal of building trust.

Aligned

AI is a cross-border, general-purpose technology. As countries formulate their domestic AI regulation, it is critical to ensure that there is alignment and coherence between national regulations and international frameworks to facilitate the wide adoption of AI tools and technologies. Increased global alignment on AI regulations, including in the context of trade, will help to facilitate the adoption, use, and interoperability of AI technologies across different jurisdictions.

Prioritize international coherence and interoperability

As a starting point, governments should look at the existing body of work undertaken by different groups on AI governance, including the Organization for Economic Co-operation and Development (OECD), the Group of Seven (G7), the Global Partnership on AI (GPAI), the United Nations (UN), and regional institutions such as the African Union (AU) and the Association of Southeast Asian Nations (ASEAN).

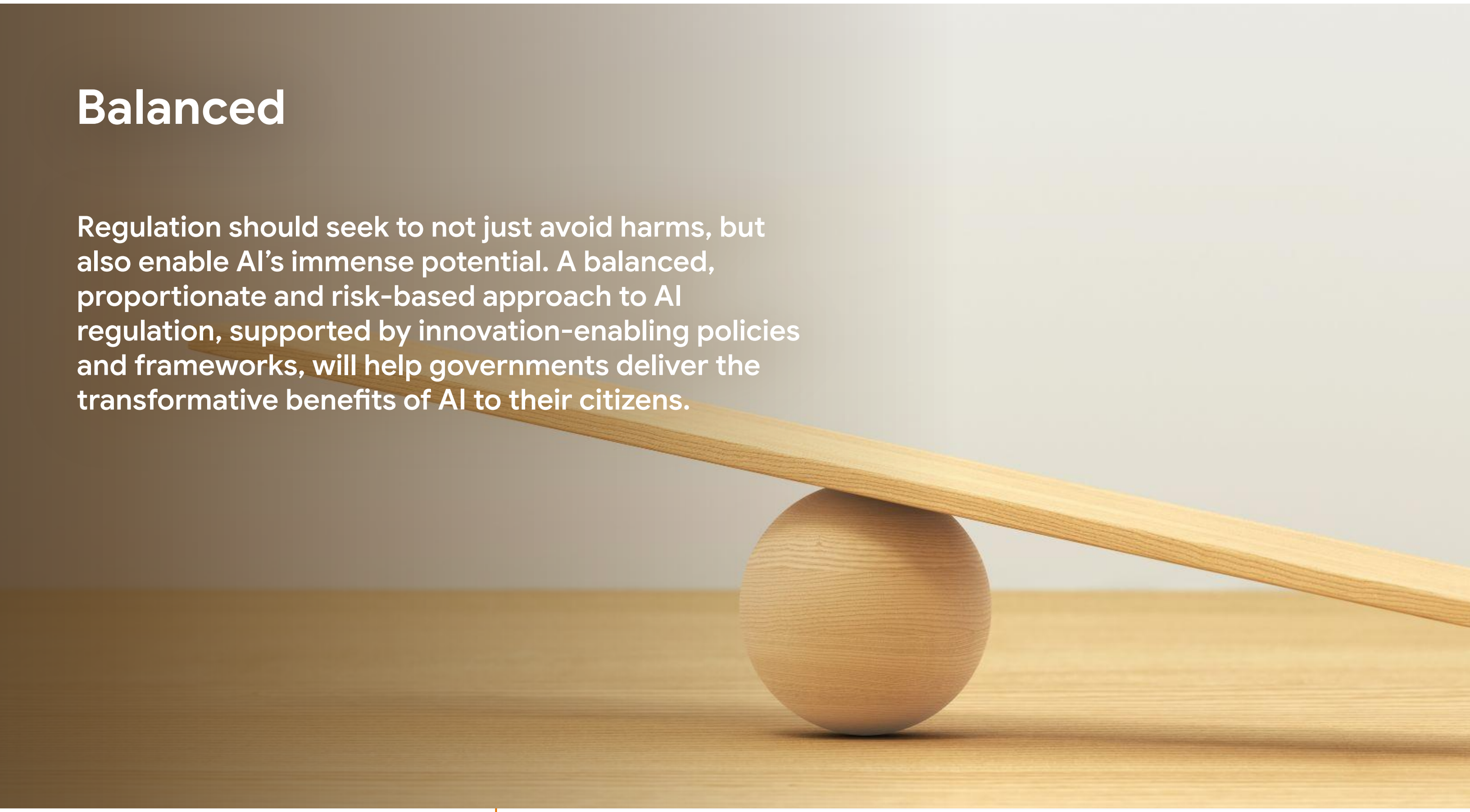
While national regulations do not have to be identical, they should be broadly consistent with the principles and frameworks set out by these international institutions and bodies. Where feasible, governments should also coordinate on any government testing requirements, adhering to the principle of home government testing, while establishing relevant information sharing mechanisms for mutual recognition amongst trusted partners.

Look to international standards and benchmarks

While this is still an evolving space, international standards bodies such as the International Organization for Standardization (ISO) and the Institute of Electrical and Electronics Engineers (IEEE) have put out international standards on AI, such as the ISO/IEC 42001, which lays out specific requirements for establishing, implementing, maintaining, and continually improving an AI management system within organizations.

Industry-driven initiatives such as MLCommons, the Frontier Model Forum (FMF), and the Partnership on AI (PAI) also provide useful reference points for governments as they consider AI benchmarks.

Governments should consider incorporating these kinds of internationally recognized standards into their domestic regulatory frameworks. These standards are based on a broad and deep foundation of expertise from a wide variety of industry and civil society perspectives. As such, they can be flexible and nimble in a way that static regulation cannot, and can easily evolve over time as technologies advance.



Balanced

Regulation should seek to not just avoid harms, but also enable AI's immense potential. A balanced, proportionate and risk-based approach to AI regulation, supported by innovation-enabling policies and frameworks, will help governments deliver the transformative benefits of AI to their citizens.

Adopt proportionate, risk-based frameworks centered on use cases

No technology, including AI, is ever risk free. But responsibly developed and deployed AI can help us make significant progress on social and economic challenges and reduce a vast array of everyday risks.

The context of use is critical in determining risk. It is vital to ensure that any regulatory framework is targeted at the applications of AI in areas with the highest risk of harm and misuse, while recognizing that these high-risk applications often also bring high-value, such as in medical applications of AI. In determining risk, governments should take into account not just the likelihood and severity of harm, but also the opportunity cost of not using AI.

Governments should also avoid designating overly broad and imprecise categories such as "healthcare" or "government/public services", or ambiguous and highly subjective domains such as "negative effects on fundamental rights and economic security" as high-risk. This will inadvertently sweep in many low- to no-risk use cases as high-risk and disincentivize the adoption of benign and beneficial uses of AI.

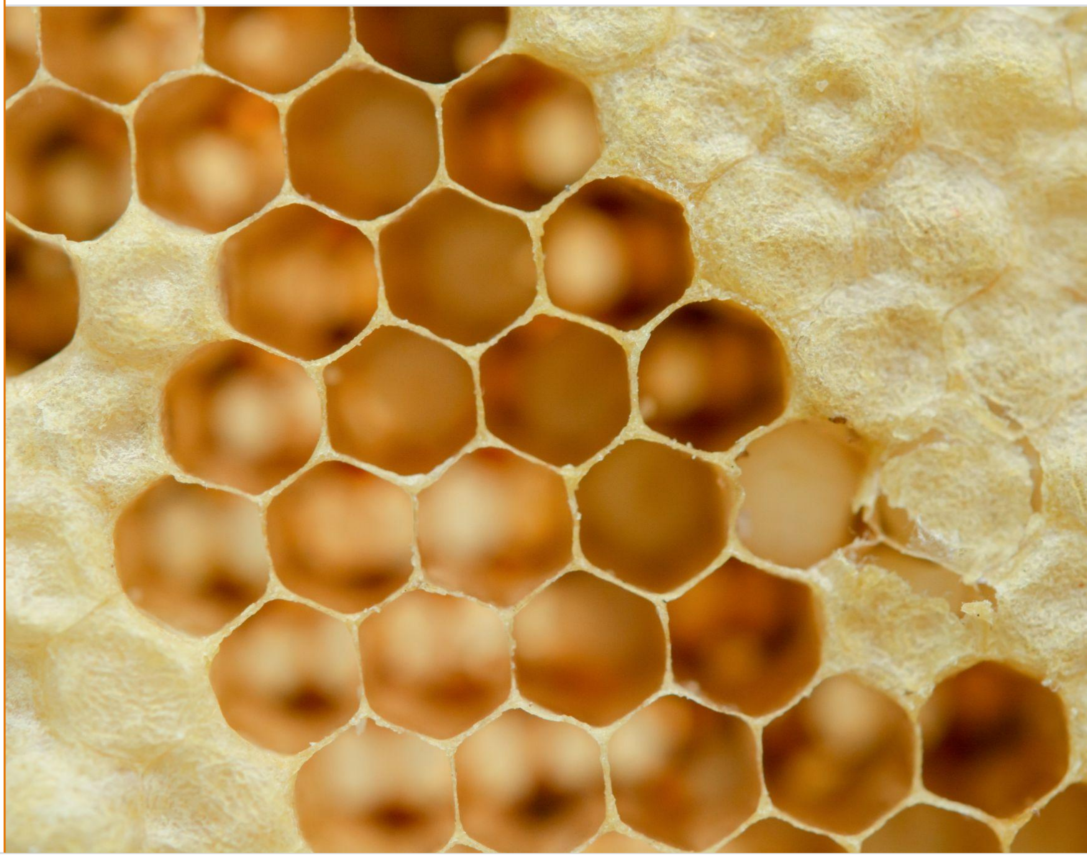
For instance, AI could be deployed for some low-risk healthcare or government functions, such as using AI to sort government archives or to book a medical appointment.

Articulate clear and differentiated obligations for the respective actors in the AI ecosystem

AI regulations should identify the key actors in the AI ecosystem — typically developers, deployers, and end users — and clearly spell out their respective roles and responsibilities. Each of these actors has a distinct governance role to play, and the actor that has control over a specific step in the AI lifecycle should bear the responsibility for that specific step.

For example, in many instances, the original developer of an AI model has little to no visibility or control as to how it may be used by a deployer, and may not have any interactions with users. Even in cases where a model is provided by the developer directly to the deployer, and no significant modifications are made, deployers will often be best placed to understand the downstream use cases and their attendant risks, implement effective risk management strategies, and conduct post-market monitoring and logging.

On the other hand, developers should be expected to provide certain information and documentation to the deployers, such as documentation of how the models were trained or mechanisms for human oversight, to enable deployers to comply with relevant regulatory requirements.



Avoid regulatory burdens for research and development, and promote access to open data to enable fair learning

It can be hard to determine precise risks in the early, iterative stages of developing AI products and services. Industry standards and bounds set by existing regulation can create clarity and enable companies to pilot innovations. Governments may also wish to consider establishing AI regulatory sandboxes to promote exploration in a lower risk environment.

Governments should enable innovative uses of openly available data which is necessary for AI development through balanced copyright rules and privacy laws. Balanced copyright rules, such as fair use and text-and-data mining exceptions, have been critical to enabling AI systems to learn from prior knowledge and publicly available data, unlocking scientific and social advances. Governments should also promote balanced privacy laws that recognize exemptions for publicly available information to enable the development of AI systems.

Ensure that transparency requirements are balanced and feasible

In designing transparency requirements, governments should take a proportionate and balanced approach, be clear about their goals, the intended audience, and the level of detail necessary to achieve these goals.

This is especially since mandated disclosures can create trade-offs with other equally important considerations such as safety, security, privacy, trade secrets, and proprietary business information. Overly onerous and granular transparency requirements can result in significant compliance burdens and costs, slowing down innovation without clear benefits.

Weigh the trade-offs between AI tools and human alternatives

When assessing whether to restrict or regulate certain AI tools, it is important to consider their performance in comparison to human alternatives. While striving for AI that is "as safe or safer" than humans is a valuable principle, it is also crucial to acknowledge that under some conditions, AI can perform tasks more accurately, reliably, or safely than the average human.

Rather than assume how AI systems or humans perform at various tasks, a practical approach is to periodically gather statistics on the accuracy, reliability, and safety of AI solutions. This comparative assessment is important because it opens the door to widely accessible and beneficial AI solutions. Moreover, even when an AI system does not outperform human capabilities, it can be sufficient to address unmet needs in certain contexts.

For example, if the average ophthalmologist diagnoses diabetic retinopathy with 80% accuracy, a rural community without an ophthalmologist could still screen for cases of preventable blindness using an AI tool with 70% accuracy. And deploying such a tool would be much more valuable than the alternative of not deploying it at all.