Regulating Before It's Too Late: Why Musk Called for Proactive AI Governance

When we talk about AI regulation, it's inevitable to envision the rise of robots taking over our world. While this concern overly relies on sci-fi presumptions and remains far from our current reality, the intangible threats AI has already raised in our world warrant serious and proactive attention. Though less scary than humanoid robots, issues like misinformation, algorithmic bias, and data misuse could culminate in profound societal impacts if left unchecked. I agree with Musk's opinion on proactive AI regulation and will explore the current risk and regulatory landscape to demonstrate why preemptive frameworks are essential before AI evolves beyond our control.

Traditionally, we allow innovation to happen first and adjust regulations after problems arise—this "trial and error" approach works for most technologies. However, in high-risk fields such as nuclear energy (Power Magazine, 2024) and biotechnology (ISAAA, 2024), humans have recognized the necessity of proactive regulation. Artificial intelligence should also fall into this category of technologies that require preventive regulation, as the consequences could be irreversible once advanced systems lose control of critical infrastructure or autonomous decision-making. In fact, different industries adopt different regulatory intensities based on their risk profiles: high-risk sectors such as the global pharmaceutical industry use rigorous pre-market testing, and agencies such as the US Food and Drug Administration require extensive clinical trials before drugs reach consumers (U.S. Food and Drug Administration, 2024). Similarly, global automotive safety standards require that cars must be certified before they can be sold (National Highway Traffic Safety Administration [NHTSA], 2023). However, in other industries where immediate human health and safety are not directly at stake, a "deploy first, regulate later" philosophy is often the norm. For example, Germany's NetzDG law requires social media platforms to remove illegal content within 24 hours of notification and imposes heavy fines for non-compliance, reflecting this

reactive approach (BBC News, 2020). This stance stems from the difficulty of predicting all potential social harms at the time of deployment and the rapid evolvement of digital platforms.

This "deploy first, regulate later" mindset can be especially concerning for AI systems that impact not just products but human perceptions. Although the harm may not manifest in physical danger, it drives subtle, cumulative distortions of how people receive and process information. These less visible risks may feel benign and are harder to anticipate, yet their consequences may be just as far-reaching and irreversible.

In the past few years where generative AI has been used to produce text, images, and other media, its capacity to blur the line between truth and falsehood has become increasingly alarming. For thousands of years, social trust has been built on a basic consensus on the authenticity of evidence. Now, as artificial intelligence gains the ability to generate convincing fake content, this foundation faces unprecedented challenges—manifesting across numerous domains, from political propaganda and fake news to corporate fraud and identity theft to the spread of misinformation on social media. Deepfake technology is eroding the basis of social trust from all angles. Research has found that our information ecosystem is exceptionally vulnerable to AI-generated content. As people are exposed to more deepfake content, their ability to distinguish between true and false becomes more difficult. Although vigilance may increase, it also increases cognitive confusion (Groh, 2021). Hancock and Bailenson's research confirms that deepfake media can be as effective as actual content in influencing public attitudes and behaviors (Hancock & Bailenson, 2021), with exposure to such content significantly altering viewers' cognitive and decision-making abilities and weakening their trust in the entire information environment. This deep-level manipulation threatens the autonomy of individual judgment and likely fundamentally undermines the basis of social consensus, exacerbates social polarization, spreads misinformation, and ultimately erodes the shared perception of reality upon which modern democracy operates.

In academic research, AI has been used to generate fake datasets and biomedical images that appear to be real but are actually fake. This has led to the retraction of papers and undermined the credibility of scientific research. Advanced natural language processing systems can generate pseudo-original content, including academic papers and literature reviews, which not only infringe intellectual property rights but also pollute the academic knowledge base with low-quality or misleading information (Chen, 2024). This contamination of the human intellectual ecosystem is like a drop of ink into a water tank. As it slowly takes over, recovering that drop of ink is almost impossible. Similarly, if we were to rely on reactive regulations, it may be too late to intervene once academic research archives become interwoven with AI-tampered content and its derivatives—especially given science's cumulative nature. While tools exist for identifying deepfakes or AI-generated content, they are neither widely available nor reliable enough to respond effectively to the rapid spread of false information (Ramel, 2023). Without forward-thinking regulations, the development of reliable detection mechanisms will always lag behind the progress of generative AI technology.

However, there are voices opposing the early implementation of strict AI regulation. Innovation-driven companies and emerging industries that rely on the profit blowout brought about by AI-powered tools believe that overly stringent regulations could stifle innovation and impede positive developments in artificial intelligence (Karnofsky, 2024). This position advocates that market forces and industry self-regulatory mechanisms can effectively address most risks, such as OpenAI's Safety and Security Committee and the Partnership on AI (OpenAI, 2024), which can respond quickly to technological changes. While these mechanisms are developed by professionals with deep technical understanding, they also have apparent limitations—the enforcement mechanisms are often dominated by large tech companies that do not represent broader societal interests and are inadequate when facing society-wide risks and challenges.

Despite the potential risks, global AI regulation is currently lax and uneven. The European Union's AI Act was approved in March 2024 but will only be implemented in phases until 2027 (Elbashir, 2024). The United States has taken a more decentralized approach, including federal initiatives such as the National AI Initiative Act and state-level regulations such as New York City's Local Law 144 (Parker, 2022), resulting in a fragmented, industry-centric regulatory landscape. China introduced rules for algorithmic recommendation and deep synthetic content management in 2022 and 2023 (Cyberspace Administration of China, 2022), but they are mainly focused on specific application areas. Other regions around the world, such as Brazil and Argentina, are exploring regulatory frameworks similar to the EU, but most are still in the early stages of discussion (IAPP, 2025). The gap between AI's growing influence and today's slow-moving regulatory systems highlights the importance of a proactive, forward-looking regulation.

To address these challenges, we need multi-stakeholder participation in AI governance. Given the complexity of AI, no single entity can effectively manage all regulatory challenges. Government, industry, academia, and civil society must collaborate to provide diverse perspectives to ensure effective regulatory policies. Furthermore, a risk-tiered approach is essential, with stricter standards for AI applications in critical areas such as healthcare or financial systems, where failures could have devastating consequences. Proactive regulation is essential not only to prevent risks but also to maintain human decision-making autonomy. As we create systems that may surpass human intelligence in some areas, humans may also face the risk of losing the ability to decide the future. Once superintelligence systems can make critical decisions, human society may find it difficult to maintain control over its own destiny, even if these systems achieve moral alignment. In the long run, the loss and transfer of decision-making power represent another significant risk that passive regulation cannot address.

Even if we decide to adopt proactive regulation, we still face unprecedented challenges. Stuart Russell argues in *Human Compatible* that controlling advanced AI systems presents unique difficulties because they may exceed their originally programmed parameters (Russell, 2019). Recent research emphasizes that current AI architectures remain fundamentally unsafe due to their strict optimization for predefined goals without understanding broader human intentions. On the other hand, as AI systems become more autonomous, the challenge of regulatory responsibility becomes increasingly complex. Who will be responsible for the decisions made by self-learning systems? The current legal framework cannot adequately answer this question. When machines make decisions beyond their original programming, the responsibility attribution becomes unclear, leaving accountability gaps that traditional reactive regulation cannot adequately address.

The impact of artificial intelligence on the learning skills of the next generation and the preservation of human essence is equally concerning. The young generation has been immersed in an algorithmic world that shapes how they process information, develop ideas, and refine cognitive abilities. As generative AI provides immediate solutions, fundamental skills such as critical thinking, deep analysis, and creative problem-solving risk atrophy (Dolan, 2025). Moreover, as machines simulate human cognition and generate complex content, we must preserve the foundations of human identity, consciousness, and moral agency as machines continuously improve in simulating human thought processes. Educational systems must adapt to this new reality, fostering students' ability to collaborate with AI while protecting and strengthening the thinking and creativity that make us unique and human. AI regulatory frameworks must also consider these broader social impacts, ensuring that technology serves human development rather than hindering our continued evolution.

Musk's observation that AI requires proactive regulation highlights an urgent need for foresight regulations and early intervention. Unlike most innovations that can learn from mistakes, artificial

intelligence has potential impacts that are profound and difficult to reverse, compelling us to adopt prevention rather than reaction as our guiding principle. This is not born from fear of technology but from clearly recognizing its powerful capabilities and potential influences. In this unprecedented field, forward-thinking is not merely wise but necessary for protecting our collective future. It may be too late if we wait until AI systems are deeply integrated into critical infrastructure and decision-making processes before taking action. Proactive regulation is not intended to stifle innovation but to create safety boundaries that allow government, industry, academia, and civil society to collaboratively guide the development of artificial intelligence. Through multi-stakeholder participation and risk-stratified regulatory strategies, we can promote technological progress while protecting society from potential harm. In conclusion, the ultimate goal of AI regulation is to guard against foreseeable risks and protect human decision-making autonomy and our core qualities as humans, ensuring that technology truly serves human society rather than becoming its master.

References:

BBC News. (2020, February 12). Social media: How do other governments regulate it? https://www.bbc.com/news/technology-47135058

Chen, Z., Chen, C., Yang, G., He, X., Chi, X., Zeng, Z., & Chen, X. (2024). Research integrity in the era of artificial intelligence: Challenges and responses. *Medicine, 103*(27), e38811. https://pmc.ncbi.nlm.nih.gov/articles/PMC11224801/

Cyberspace Administration of China. (2022, January 4). *Cyberspace governance policy document*. China Cyberspace. https://www.cac.gov.cn/2022-01/04/c_1642894606364259.htm

Dolan, E. W. (2025, March 21). AI tools may weaken critical thinking skills by encouraging cognitive offloading, study suggests. *PsyPost.* https://www.psypost.org/ai-tools-may-weaken-critical-thinking-skills-by-encouraging-cognitive-offloading-study-suggests/

Elbashir, M. (2024, April 22). *EU AI Act sets the stage for global AI governance: Implications for US companies and policymakers.* Atlantic Council: GeoTech Cues. https://www.atlanticcouncil.org/blogs/geotech-cues/eu-ai-act-sets-the-stage-for-global-ai-governance-implications-for-us-companies-and-policymakers/

Groh, M., Epstein, Z., Firestone, C., & Picard, R. (2021). Deepfake detection by human crowds, machines, and machine-informed crowds. *Proceedings of the National Academy of Sciences,* 119(1), e2110013119. https://doi.org/10.1073/pnas.2110013119

Hancock, J. T., & Bailenson, J. N. (2021). The social impact of deepfakes. *Cyberpsychology, Behavior, and Social Networking, 24*(3), 149–152. https://doi.org/10.1089/cyber.2021.29208.jth

International Association of Privacy Professionals. (2025, January). *Global legislative predictions 2025* (J. Bryant, Ed.). IAPP. https://iapp.org/resources/article/global-legislative-predictions/

ISAAA. (2024, May 15). *EPA, FDA, and USDA issue updates to Joint Regulatory Plan for Biotechnology.* Crop Biotech Update. https://www.isaaa.org/kc/cropbiotechupdate/article/default.asp?ID=20814

Karnofsky, H. (2024, December 16). *Developing AI risk management with the same ambition and urgency as AI products.* Carnegie Endowment for International Peace. https://carnegieendowment.org/research/2024/12/developing-ai-risk-management-with-the-same-ambition-and-urgency-as-ai-products?lang=en

National Highway Traffic Safety Administration. (2023). *Vehicle certification process* (Report No. MC-10235486-0001). https://static.nhtsa.gov/odi/tsbs/2023/MC-10235486-0001.pdf

OpenAI. (2024, May 28). *OpenAI board forms safety and security committee.* OpenAI. https://openai.com/index/openai-board-forms-safety-and-security-committee/

Parker, L. (2022, June 29). *National Artificial Intelligence Initiative.* U.S. Patent and Trademark Office. https://www.uspto.gov/sites/default/files/documents/National-Artificial-Intelligence-Initiative-Overview.pdf

Power Magazine. (2024, June 20). *The ADVANCE Act: Legislation crucial for a U.S. nuclear renaissance clears congress — Here's a detailed breakdown.* POWER Magazine. https://www.powermag.com/the-advance-act-legislation-crucial-for-a-u-s-nuclear-renaissance-clears-congress-heres-a-detailed-breakdown/

Ramel, D. (2023, July 10). *Researchers: Tools to detect AI-generated content just don't work. VirtualizationReview.com*. https://virtualizationreview.com/articles/2023/07/10/ai-detection.aspx

Russell, S. (2019). *Human compatible: Artificial intelligence and the problem of control*. Viking.

U.S. Food and Drug Administration. (2024). *Drug approval process.* https://www.drugs.com/fda-approval-process.html