Center for AI Policy

From Specific to Systemic: Broadening AI Regulation Beyond Use Case

Executive Summary

- Influential voices like NVIDIA and IBM[1] have suggested regulating AI based on specific use cases[2] and asking existing regulators to oversee AI being used in each industry, with airline regulators tackling airline AI, medical regulators tackling medical AI, etc. This method fails to address the unique risks inherent in new general-purpose AIs (GPAIs) like GPT-4, namely: misuse across a broad array of use cases, unprecedented rapid progress, and rogue systems that evade control.
- To properly address these risks and keep the American public safe, we need to establish a central regulator which will:
  - Reduce government waste and needless redundancies
  - Bring leadership necessary for coordination
  - Facilitate effective, risk-focused, pre-deployment regulation
  - Introduce much-needed proactivity into AI regulation
  - Account for new capabilities that fall outside existing regulators
- One promising framework for a central regulator is a tiered approach that categorizes models according to indicators of capabilities, and scales regulatory burden with capabilities.

[1] For example, **Jim Fan (NVIDIA):** "The right place to regulate AI is at the APPLICATION layer…adding burdens to foundation model development unnecessarily slows down AI's progress" and **Christina Montgomery (IBM):** "IBM urges Congress to adopt a precision regulation approach to AI. This means establishing rules to govern the deployment of AI in specific use cases, not regulating the technology itself". **But also note** that there is disagreement in industry too, as other major industry players, such as **Microsoft**, have called for regulating beyond the application layer.

[2] AKA the "application layer" or "vertical regulation" or "precision regulation."

What has changed?

Regulating by use case made sense as recently as 5 years ago, when essentially all AIs were tailored to narrow circumstances and unable to accomplish tasks outside those circumstances. When AIs were narrowly tailored, we could manage AI risk well by identifying the riskiest use cases and holding AI to higher standards in those domains. However, today's general-purpose AIs (GPAIs) are importantly different from the narrowly tailored AIs of the past and pose three unique challenges which must be accounted for with general-purpose regulation.

Why are GPAI's challenges unique?

Misuse occurs across a broad range of use cases. The general nature of GPAIs carries an inherent risk of misuse, even if the use case they are ostensibly developed for isn't particularly risky. For example, a GPAI that was intended to write travel guides could instead be used to generate pornographic stories about children. Many misuses are already possible with the GPAIs we have today, like crafting convincing phishing emails. Others will likely become more of a concern as GPAI improves, such as exploiting cybersecurity vulnerabilities, or developing chemical and biological weapons, the latter potentially only two or three years away. And these are just the ones we know about, further development is likely to create totally new capacities that will present new forms of risk we've yet to encounter.

Progress is accelerating faster than we can keep up with. GPAI capabilities are advancing quicker than many thought they would, and the rate of advance will likely get even faster over time. This is due to self-reinforcing trends in GPAI development, where improvements in AI capabilities are used for better developing future AIs. Improved AI systems could also accelerate progress in other areas, leading to a feedback loop of idea generation (e.g. more ideas → more economic output → more AIs → more ideas) which could cause unprecedented, rapid change across the economy.

Advanced GPAI could go rogue. Current research is being aimed at developing GPAI that is smarter than us, a process that runs the risk of creating AIs that can act on their own, with no human in the loop[3]. Autonomous AIs might be good; they might pursue goals that we approve of and that align with what we value in the world. But they might

[3] This might come about naturally from current methods for developing AI, but some people are also pursuing this directly, including large tech companies like Nvidia, DeepMind and OpenAI.

also be bad, as our lack of technical guardrails leaves us vulnerable to the risk of creating rogue AIs with goals that go against our own, which could cause massive damage to society at large. A single company with a profit motive to release their product as quickly as possible should not be allowed to unilaterally decide that their GPAI is "guaranteed" to lead to a better future. Instead, nations and experts from safety, industry and academia agree that we should set minimum safety standards so that all GPAI developers will be expected to develop with caution.

Why do we need a central regulator?

It will reduce redundancies and wasteful government spending. Under use-case regulation, a single GPAI system might require approval from nearly all sector-specific regulators, and much of this work will be duplicated analysis that a single GPAI regulator could do more efficiently. A central regulator offers a different vision, one of efficiency and coordination where work is shared rather than repeated, requiring less people to get the same work done. Beyond increasing efficiency, establishing a central regulator would also be essential in making sure AI talent in government is not spread too thin. With serious competition for top talent, the government is far from matching the monetary benefits of industry, so allocating talent efficiently will be imperative for effective regulation.

Use-case regulation suffers from a lack of leadership. We've already been testing regulation by use case, and it's failing. A 2022 study on compliance with existing AI regulation[4] found a "weak and inconsistent" implementation of legal requirements across federal agencies, most failing to become compliant even years later. This is likely due to a pitfall inherent to (exclusive) use-case regulation: coordination problems borne out of a lack of leadership, where it's not clear who is responsible for holding agencies accountable. In this sense, current regulation is like trying to run an orchestra without knowing who the conductor is. Until we have a central regulator to orchestrate the multitude of agencies on GPAI, our regulation will be in a similar state of disarray.

Pre-deployment regulation is crucial. Advocates for use-case regulation tend to support pre-deployment regulation of AI only in high-risk environments like healthcare, resulting in regulation for AI used to interpret MRIs but nothing for GPT-4. Some take it even farther, arguing we actually can't tell if a GPAI is going to be harmful until we

[4] See the AI in Government Act of 2020, Executive Order 13,859 on AI Leadership, and Executive Order 13,960 on AI in Government.

deploy it, and thus should stay away entirely from regulating GPAI developers. This seems to totally ignore the progress currently being made on pre-deployment safety efforts like red teaming[5] and model evaluations, which can help reduce future misuse of the model by actively probing for undiscovered dangerous capabilities beforehand. It also fails to account for the fact that capable models create risks even before they are deployed. Never totally secure in the hands of developers, they are at constant risk of an accidental leak or theft by malicious adversaries who can then exploit the model capabilities already present pre-deployment.

Reactivity is insufficient. A strand that runs through properly addressing the risks from GPAI is the need for proactive regulation on top of current reactive regulation. Proactive regulation allows us to safeguard the public before a major harm. Without it, we'd be left to confront explosive change and increases in capabilities that would likely outpace our slow lawmaking process, leaving us vulnerable for the months or years it takes to form appropriate regulation. We'd also have no way to properly address the catastrophic and irreversible risks that might come from something like a rogue AI which we've lost control of, where the only chance to reduce the risk comes from proactive safety measures.

Lack of clear responsibility for new capabilities. Use cases that don't fall neatly under a current regulatory body's purview will likely fall through the cracks, because it's unclear who would regulate the uses that don't neatly fit into the existing regulatory scheme. Our current system could sort this out eventually, but a slow bureaucratic process likely spanning months (or years) is not fit to address a technology developing at such a breakneck pace. To address that pace we need to proactively develop a process that affords clarity and speed. We need a central regulator that can catch what falls through and thoughtfully decide how these uses can be best addressed, whether that be initiating regulation or delegating the regulation to an appropriate use-case regulator.

What should happen?

We believe that **GPAIs with risky capabilities should be classified as high-risk even if they are targeted at a low-risk sector,** governed by **a central regulator that sorts**

[5] Generally testing models as if you were an adversary or malicious actor to see if you can get them to produce harm (read here for further context).

the risky AIs from the safer ones and prevents risks at every stage of development, not just the application stage.

As a start to this method, we think that **risk-based regulation should focus on indicators that correspond to more powerful GPAIs.** We might look at a range of things like computational power, parameter count, cost of training, or benchmark performance[6] to assess which AIs are highly capable, and thus should be regulated. This could inform a targeted approach where we regulate the most powerful GPAI models at different stages or levels of development, such as hardware (e.g. large supercomputers), creation (e.g. lengthy training runs), deployment (e.g. ChatGPT), and possession (e.g. access to weights).

The entire AI industry might be too large to comfortably regulate out of a single office, but we can and should have centralized regulation for the particular part of the AI industry that is focused on developing advanced, general-purpose AI. We urgently need a central regulator for GPAIs to **tackle catastrophic risks from AI head-on and address them at their root cause.**

[6] Benchmarks are essentially standardized tests that measure the performance of AI systems on specific tasks and goals.