

BROOKINGS

RESEARCH

Balancing market innovation incentives and regulation in AI: Challenges and opportunities

Kevin A. Bryan and Florenta Teodoridis

September 24, 2024

Key takeaways

- Regulating new, fast-changing technology is very difficult, both in theory and empirically
- The direction of technology is determined not only by regulation but also by market incentives
- Innovation is usually under-incentivized, and too-strong regulation on AI exacerbates underinvestment and reduces experimentation

How can society capture the benefits of artificial intelligence (AI) while minimizing fraud, safety concerns, and unemployment that AI could produce? [Some AI experts argue](#) that regulations might be premature given the technology's early state, while [others believe they must be implemented](#) immediately to ensure AI systems are developed and deployed responsibly. Central to this debate are two implicit assumptions: that regulation rather than market forces primarily drive innovation outcomes and that AI should be regulated in the same way as other potentially harmful products which are more fully developed.

Both assumptions are incorrect. When and how the market distorts the direction of technological innovation in the presence of externalities and uncertainty and when regulation is useful or harmful, are topics that have long been studied in contexts

outside of AI. Promoting socially beneficial AI depends not just on technical and legal knowledge but on lessons from economics and management in how the trajectories of new technologies unfold.

Concern over market forces

The primary concern of innovation policy is that laissez faire markets may not provide the necessary incentives to ensure the socially optimal rate and direction of research. Early inventors make possible later breakthroughs whose profit they do not capture and that early research often involves a fixed cost paid only by the initial researcher; it is easier to copy or build on a known invention than to create it in the first place. When there are multiple paths on which a technology can develop, we want to simultaneously encourage firms to spend on research in the first place, ensure that they work on the most socially beneficial technologies ([rather than just race to market](#) [↗](#) with harmful or shoddy ones), and [guarantee a broad diversity](#) [↗](#) of research options if switching research directions is hard.

The fundamental problem for a regulator is therefore clear. They need to incentivize firms to do more research because inventions create positive spillovers for follow-on inventors, but simultaneously they need to dull incentives for inventors to work on harmful technology, mediocre technology with few spillovers, and technology overly focused on a single development path or hyped area. And the regulator needs to achieve this balance despite facing great uncertainty about harms and benefits.

To illustrate these concerns, let us examine an argument by the economist [Daron Acemoglu](#) [↗](#). Given that generative AI, especially for “foundation models,” is often very costly to create, he argues the path of technology will be heavily influenced by large, resource-rich companies. This path could in many cases lead to excessive inequality through automation-driven job displacement; market concentration that stifles competition and limits innovation from smaller entities that cannot compete on the same scale; and political and social harms through the spread of misinformation and polarizing content which undermines public discourse and democratic processes. That is, firms will disproportionately innovate to save on labor costs because they do not account for the social benefit of worker wages; will become too concentrated because the high fixed costs of AI development mean only a few firms can compete;

and will underweight safety risks to society and democracy because these risks do not bear directly on their profit.

This paints a pessimistic picture regarding the role of markets in developing AI. That said, the future of this technology is not determined by the free market alone. A primary motivation for ensuring a thriving academic research sector that is free from commercial incentives has been to [mitigate the limitations](#) of free markets by ensuring experimentation across broad research trajectories. Vannevar Bush, in his seminal work ["Science, The Endless Frontier,"](#) emphasized the importance of support for scientific research that is free from commercialization goals to drive innovation and societal progress.

Shifting AI development toward academia is, however, no panacea. Academic scientists also respond to incentives. [Scientists target](#) certain journals and not others, [consider reputational benefits](#) when selecting their projects, and evaluate the likelihood of obtaining grants if pursuing one project or another. The impact of these considerations on project diversity is exacerbated by [the expansion of the knowledge frontier](#) that forces researchers to specialize in increasingly narrower niches and hence depend more heavily on [collaboration](#) across technological niches. As a result, the breadth of potential trajectories a scientist would be aware of or even consider pursuing narrows over time, absent intervention that tries to incentivize diverse pursuits. External factors such as funding conditions and [costs of research tools](#), market [demand-pull factors](#), and features of local research environments, including geographic endowments and firm policies, can also affect researchers' choice of projects. These factors may increase research productivity and diversity by reaching scholars from different areas and enabling broad experimentation. However, it may also limit the breadth of explored research trajectories by incentivizing a focus on particular research or on solving specific questions.

Is there a way to balance the advantages and disadvantages of for-profit firm and academic innovation? Some areas of technology development benefit from the focusing power of the profit motive, while others are easier to motivate in alternative environments like an academic lab. Thus, broad innovations like AI often benefit by drawing on [complementarities](#) between firm and academic expertise and resources. These complementarities are essential during the [early stages of an industry development](#), and particularly so when the [technology is general purpose or](#)

[enabling, like AI ↗](#), meaning its value creation potential depends on repeated cycles of innovation that involve diverse economic actors from both producing and application sectors.

For example, the [development of quantum computers ↗](#), another contemporaneous enabling technology, accelerated after both firms and academia became involved and engaged in collaboration. A wide variety of technological paths were investigated by these diverse researchers. By combining the benefits of academic inquiry less affected by commercialization concerns with the firm advantage in access to costly resources, quantum computing, thus far, has not been pushed into narrower trajectories in search for immediate, socially-suboptimal returns to investment.

[Several experts and companies ↗](#) have discussed the benefits of promoting open-source practices in AI as an approach to taking advantage of the complementarity in skills and resources across sectors. The hope is to achieve experimentation in diverse trajectories by tapping into a diverse set of researchers who can share their knowledge and hold each other accountable, by reducing barriers to entry for startups and other economic actors that want to contribute, and by enabling the public to monitor and influence AI development trajectories that increase social welfare.

Note, however, that coordination between firms and academia through open-source practices still depends on their respective incentives. While open sharing is a foundational principle of universities, it is not one of firms. Firms need to [anticipate enough value capture ↗](#) from their open innovation efforts to engage. For general purpose or enabling technologies, this is [particularly difficult to achieve ↗](#) as it depends on a [variety of factors ↗](#) such as in-depth complementary knowledge about potential applications alongside other complementary assets, a strong intellectually property protection regime, and clarity about a dominant design. These are difficult to achieve when technological uncertainty is high.

Concerns over regulation

A third player who can influence the future trajectory of AI outcomes is, of course, government. Those who argue about the benefits of regulation in limiting market concerns generally focus on regulating the output of AI innovation efforts, not the

direct process of innovation itself. The goal is to prevent AI developments that cause harm. However, regulating the output of uncertain and rapidly developing technology is necessarily different from traditional product regulation.

When innovation can cause harms, [regulators have three basic options ↗](#): ex ante restrictions such as bans on use or further research, ex post withdrawal from the market, or liability for harms. In principle, these policies can all achieve the same goal of aligning market incentives with socially desirable outcomes. However, their relative efficacy depends on regulators' knowledge about technology's potential risks and benefits at any given time.

To make clear how serious the informational problem is for regulators, consider the earliest proposed laws related to AI. The [European Union Artificial Intelligence Act \(AI Act\) ↗](#) was proposed in early 2021. This [initial EU proposal ↗](#) does not, across over 100 pages, use the words "large language model," "LLM," "transformer," or "generative" a single time. The definition of high-risk AI systems used in the initial EU AI Act includes those used in education and law enforcement but it does not in any way constrain independently-acting agent systems. Early U.S. state-level laws likewise focus on concerns of a far narrower scope than what is possible in 2024. For instance, [California's BOT Act ↗](#) of 2018 restricts AI communication related to sales or elections without disclosure that an AI is being used. Illinois' first AI-related law ([AI Video Interview Act of 2019 ↗](#)) requires consent for AI evaluation of first-round interviews, including a bias audit, and New York's first AI rule ([Local Law 144 ↗](#)) was focused on requiring bias audits when AI is used in hiring.

Why are these regulations so disconnected from present day AI safety concerns? The primary worry from AI three years ago, in the view of policymakers, was its use either to further discrimination or to permit anti-liberal surveillance of citizens ([called "AI-tocracy" by some scholars ↗](#)). Because the legislative process takes time, and politicians are not omniscient, regulation is often tailored to technological threats or worries that may ex post appear minor.

Moreover, regulation may limit otherwise useful technological development trajectories that would mitigate the very harm being considered. As an example, consider alignment, the problem of getting an AI system to do what users intend, or interpretation, the problem of understanding why an AI system does what it does; on

both counts, we have [growing evidence ↗](#) that more complex models may be [easier, not harder, to align ↗](#) and interpret. Regulation that intends to prevent risky AI by limiting the size of the model may, therefore, inadvertently prevent the development of the very technology that would solve that problem.

Things are worse yet when regulatory uncertainty intersects with the need for inventors to experiment to reduce technological uncertainty. Because the regulator might not observe everything firms and scientists know, it may shut down useful innovation too early or inadvertently permit potentially harmful innovation to progress. Liability makes the firm responsible for many of the downside risks of their innovation, even as they do not fully capture the upside of useful experimentation; the social value of a major AI breakthrough does not accrue fully to the inventor, hence when liability is too strong, firms innovate too little. Likewise, an incentive system that heavily incentivizes AI breakthroughs without liability for harm induces racing behavior among firms who do not fully bear the cost of downside risks which they may be aware of but regulators did not foresee. [Firms will race to get to market first ↗](#), regardless of safety risk, because they get the upside of first-mover advantage in the market without the downside should their invention prove dangerous. It is not an easy task for a regulator to know on which side of that ledger the power of law ought to apply.

For example, imagine there are two ways to build AI models, of which the benefits and harms are both initially unknown. A breakthrough happens in quality for the first model, but there are clear harms which become visible. The innovation economist Joshua Gans [points out ↗](#) that you may want to continue work on the model known to be harmful simply as a matter of costs versus benefits. Because this model has also already shown benefits, and the other model is still uncertain on both the cost and benefits sides of the ledger, continued development of the partly-harmful model may make it more likely, not less, that benefits exceed harms. Worse yet, punishing firms through liability for continuing development of that model may push them into working on a model that is still potentially dangerous but whose benefits are also more limited in expectation.

Artificial intelligence is far from the first technology which is simultaneously potentially transformative and potentially very dangerous. Consider similar lessons from the history of nuclear power. The social benefit potential for nuclear power was clear by the end of World War II, as was the enormous destructive potential of splitting the

atom. How should firms like General Electric and Westinghouse be permitted to develop nuclear energy in a safe way, giving us the benefits of clean “too cheap to meter” electricity while avoiding a global thermonuclear cataclysm?

Energy regulators in the 1950s were operating under a veil of tremendous uncertainty. Dozens of different nuclear reactor types with wildly different safety and efficiency profiles were being [actively researched](#) ⁷. What fissile material and coolant should be used? Should reactors “breed” fissile material or refuel? How and when do reactors need to be “scrammed” for safety? How should we develop and share information about materials used in the plant to ensure safe best practices spread? What aspects require new laws and what is instead already covered by existing energy regulation and traditional firm liability? How do we prevent firms from “racing” to develop saleable but unsafe reactors?

As with AI regulation today, what we saw historically was a mixture of policies encouraging innovation with one hand and raising costs with the other hand. For example, the cost of nuclear power development was reduced via R&D bond subsidies and subsidized liability insurance in the [Price-Anderson Act](#) ⁸, while the cost and regulatory risk of developing alternative nuclear technologies which were not already advanced by the late 1950s increased. Indeed, despite highly varied research programs in the 1950s, by 1962, the Atomic Energy Commission began accepting only applications from “proven reactor concepts,” while safety regulation tilted toward [precisely those types of reactors](#) ⁹. As a result, global nuclear energy became dominated by light water reactors which even contemporaneously were not viewed as the most promising design either for efficiency or safety. Regulators mistakenly set rules on the basis of contemporaneous evaluations of harms and benefits, without considering their imperfect information impact on experimentation in a broad set of research trajectories. In 2024, we are only now recovering ideas like molten salt reactors which were stunted by 1960s regulation.

Potential paths forward

Given these challenges, what can be done? We suggest that jurisdictions considering regulating AI consider four factors.

First, policymakers should consider the process of innovation, not only the outcome, by incentivizing collaboration and coordination between complementary firm and university innovation which balances the advantages and disadvantages of each, as discussed above. Certain aspects of AI such as safety worries, basic advances which are hard for individual companies to profit from, and technology features to ensure AI is a complement rather than a substitute for human labor may be easier to incentivize outside the private sector. Regulation that implicitly limits cross-organization partnerships, such as [data privacy & restrictions >](#), can be harmful. Academic-industry collaboration can be valuable to generate useful information about the cost and benefits of particular AI developments, where that information can help better target future regulation.

Second, regulators focused on the ensuring responsible AI outcomes should specify the precise market failures they believe exist and specify why market incentives make them worse. Some regulatory questions are straightforward: For example, when a factory pollutes a river, the government can tax the externality. On the other hand, regulation of innovative industries where the nature and magnitude of the externality is changing over time as the technology develops require clarity from regulators about what exactly is being regulated in order to understand whether a market failure exists. Consider a regulator concerned that AI models which can't be controlled precisely are dangerous. It may therefore seem that "uncontrolled AI" is a market failure which developers will not internalize, hence we require taxes, bans, or liability rules to ensure safety. However, the potential for controlling AI develops in real time both in response to regulation and to market competition among innovators. For instance, more controllable AI may develop as a byproduct of attempts by competing firms to outcompete existing foundation models on tasks like code-writing assistance, even if the profit incentive here has nothing to do with control for safety purposes. Restrictions on innovation that are meant to indirectly solve a currently-existing market failure may therefore make things worse.

A similar issue comes up when attempting to regulate "moral" behavior. The Nobel Prize winner [Jean Tirole and his coauthor Mathias Dewatripont >](#) consider a general case of competing firms who care both about profit and moral issues, consumers who prefer to buy the best product for the price, and partially substitutable products which are either safer but slightly worse or more harmful but slightly better. More intense competition means it is harder for firms who want to act morally to make a sale. On the other hand, that same intense competition also means best-case profits are not that

high, hence “acting morally” is less costly. In the AI case, then, it is a priori unclear whether social welfare is higher in a very competitive open market for firms, perhaps driven by open source, or a highly regulated one where only very few firms can even attempt to advance AI. The link between proposed regulation in AI and its eventual effects is therefore very difficult to foresee, and hence regulation focused on goals rather than on methods of achieving those goals is more likely to succeed.

Third, policymakers should consider how regulation under technological uncertainty is different from standard product liability considerations. The regulator cannot know perfectly either what the benefits and harms of continuing development of already existing technology will be nor can they know the relative benefits and harms of alternative technological trajectories scientists would shift to following a ban. The existing theoretical literature is substantially more supportive of ex post liability considerations rather than ex ante bans or restrictions given the uncertainty of the different potential trajectories. Only specific, foreseeable, and preventable harms should involve ex ante restriction on AI developers. There is real danger in regulating away the experimentation that may itself solve tricky social problems or in regulation restraining the development of AI such that, as with nuclear power, we remain tied to a technological path that is actually worse for both efficiency and safety.

Finally, regulation of quickly evolving technological outcomes needs to be itself nimble and modifiable. Note again that none of the 2021 vintage regulations considered in the U.S. or EU had a word to say about LLMs or generative AI, which has become the primary regulatory concern in 2024. An AI regulation set today without flexibility is unlikely to foresee either the harms or benefits of AI as it will exist in 2027, let alone 2037. For example, the exact same [“ImageNet ↗”](#) image recognition breakthroughs that allow widespread human surveillance are also critical to the development of self-driving cars. As a strategy for such cases, [Acemoglu and Lensman ↗](#) argue theoretically that when harms are correlated and irreversible, and when Pigouvian taxes cannot be tailored by industry, a planner may want to focus on limiting AI adoption in high-risk sectors until less harmful sectors have proven those harms are unlikely. That is, AI adoption inside of a nuclear plant ought to be avoided until controllability is proven in industries with less obvious negative externalities. We recognize that different jurisdictions vary in their willingness or ability to modify existing regulation as technology changes. Standing committees such as the proposed [AI Safety Institute in the United Kingdom ↗](#) may help here, though only if they are seen by stakeholders as representing fair attempts to balance harms and benefits under

uncertainty rather than being captured by regulated firms or groups with idiosyncratic preferences.


Conclusion

Regulating rapidly developing technology requires a different model from traditional policy. Why? First, the direction of technology is uncertain, and hence premature regulation risks cutting off the experimentation from investigating a broad set of trajectories which may self-solve for the potential harms. Second, the direction of technology development is a function of market forces, academic science, and regulatory nudges, not regulation alone. The nature of optimal AI regulation therefore requires understanding not just whether AI in its current state can be helpful or harmful but rather whether that balance is better handled by regulation or market experimentation, or whether regulation should be used to shut research down versus permitting continued development in various trajectories.


We do not object to the role of regulation in ameliorating harms. Social and environmental goals often conflict with private sector incentives. Indeed, as [Joseph Schumpeter](#) and [Kenneth Arrow](#) taught us, innovation is an area where laissez faire markets are particularly unlikely to provide optimal incentives. The question is not whether to regulate but rather to what extent the harms of AI are best solved by market forces and when they are best solved by regulators who themselves have imperfect knowledge. AI development has the potential to be an [epoch-defining technological change](#). It also has the potential to impose substantial costs on humanity. The balance between market-driven innovation and regulatory intervention remains crucial, as we strive to harness the transformative potential of AI while mitigating its risks and ensuring that its benefits are equitably distributed across society.

AUTHORS



Kevin A. Bryan Associate Professor of Strategic Management -
University of Toronto  @Afinetheorem



Florenta Teodoridis Associate Professor of Management and
Organization - USC Marshall School of Business  @florentaT

Copyright 2024 The Brookings Institution