RESEARCH MEMO
DECEMBER 2024

International Governance through Domestic Law in the Forthcoming UK Frontier AI Bill

Authors: Nicholas A. Caputo, Haydn Belfield, Jakob Mökander, Matteo Pistillo, Huw Roberts, Sophie Williams and Robert Trager

International Governance through Domestic Law in the Forthcoming UK Frontier AI Bill

Nicholas A. Caputo,*[1] Haydn Belfield,[2] Jakob Mökander,[3] Matteo Pistillo,[4] Huw Roberts,[5]
Sophie Williams,[6] Robert Trager**[1]

* Corresponding author: Nicholas A. Caputo nick.caputo@oxfordmartin.ox.ac.uk
** Senior author
[1] Oxford Martin AI Governance Initiative, University of Oxford
[2] Centre for the Study of Existential Risks, University of Cambridge
[3] Tony Blair Institute
[4] Apollo Research
[5] Oxford Internet Institute
[6] Centre for the Governance of AI

Given this document's scope, inclusion as an author does not necessarily entail endorsements of all aspects of the report.

Executive summary

The government of the United Kingdom is currently in the process of developing a bill to regulate frontier AI. Such a bill must have an international scope because the companies seeking to create these systems are scattered around the world and because AI models and the effects that they cause travel easily across borders. But the international implications of frontier AI regulation have been relatively neglected in discourse about the bill so far.

To address this neglect, the Oxford Martin AI Governance Initiative recently convened a group of experts to explore how the forthcoming UK frontier AI bill can be best shaped to achieve the UK government's goals of effective regulation while remaining narrow and pro-innovation. The core takeaways from the convening are as follows:

1. **The United Kingdom should act now to secure a position of leadership in frontier AI**: The field of international regulation of frontier AI remains relatively empty. A strong and well-designed bill that sets the model for frontier regulation would put the United Kingdom in a leadership position to shape further developments in the area. Furthermore, a clear bill would both improve safety and clear the way for innovation and economic growth by providing predictable rules that facilitate compliance on the part of AI developers.
2. **Domestic law is a key part of international regulation**: While direct regulation of foreign AI companies and other entities is likely a necessary part of ensuring the safe development of frontier AI, well-designed domestic laws can shape activity abroad without necessarily raising hard issues of extraterritoriality and the like. Mechanisms like a 'London Effect' and modeling best practices allow domestic law to influence foreign actors. As such, designing domestic laws with international effects in mind would let the government maximize its regulatory effectiveness.
3. **The government must balance expanding its own reach and its reliance on others**: Frontier AI regulation must cover foreign entities, but the government should not go too far in trying to assert domestic power outside its own jurisdiction. An international system of evaluators and regulatory authorities that provide mutual safety assurances across jurisdictions would help resolve this dilemma, but no one state can create such a system alone. As such, the government should shape the bill to rely on credible assurances from foreign regulatory authorities where possible while retaining the power to prevent harm directly in emergencies. A clause to the effect that the law would apply to companies whose systems could have 'substantial and foreseeable harmful effects' on the United Kingdom would help provide this backstop while also keeping it constrained.

4. **Robust international regulation promotes the domestic economy**: Much of the United Kingdom's economic advantage in AI will likely come from specialized AI development and the deployment of models in particular use cases. Regulation at the frontier level would enable downstream users to deploy frontier models without worrying about safety risks and focus the regulatory burden on international frontier companies rather than on smaller domestic start-ups and other enterprises.
5. **The UK AI Safety Institute (AISI) should continue to advance the state of the science as an arm's-length body (ALB)**: UK AISI is the world leader in frontier model evaluations and has made strides in the science of AI safety and in coordinating with companies and with other evaluators. This cooperative approach has been effective so far and illustrates the value of making the United Kingdom the best place to do business and work with regulators. Turning AISI into an ALB will allow it to continue exerting influence to achieve the government's goals. How to supplement AISI in its scientific role with direct regulation (for example, by creating an independent frontier regulator or by expanding the responsibilities of existing bodies) is outside the scope of this report but is a key question for lawmakers.
6. **The government should emphasize offering free evaluations and safety certification to open-weight AI developers to incentivize participation in the regulatory regime**: Open-source models are generally transparent, decentralized, and accessible to consumers and companies that cannot afford proprietary models. However, they can be difficult to regulate under a frontier framework because their harms cannot be as easily attributed to the model provider that would otherwise be subject to regulation. Offering free evaluations and similar safety tools to open-source developers would help bring them into the frontier regulation regime without imposing excessive costs on these groups.

The convening focused on six core means of international regulation of frontier AI, each with its benefits and costs. Different ways of writing the frontier AI bill will make use of various methods of regulation, and lawmakers should carefully consider how to accomplish their goals of safety and growth with these tools.

Table 1: Benefits and costs of core means of international regulation

Means of regulation	Benefits	Costs
Extraterritorial regulation	· Most direct control over content and implementation of regulations · Could be key backstop in case of emergencies	· Controversial and serious expansion of power unless constrained · Limited efficacy without cooperation from abroad
De facto London Effect	· Relatively direct way for domestic regulation to shape international companies without overreach · Leverages expertise and talent of UK regulators	· Relies on compliance by frontier companies instead of forking or avoidance · Not fully predictable which rules will diffuse and how · May require more specific domestic rules than optimal in changing technical environment
Multilateral agreements	· Leadership here will set framework for future treaties · Creates framework for enforcing requirements against other states	· Likely to take a long time to come into being · Hard to get all necessary states to cooperate with each other to create treaty · Difficult to monitor especially over long-term
International harmonization	· Builds on existing strengths with AISI and similar bodies · Likely key to complete coverage of frontier AI · Expert-driven and technical process	· Requires willing and capable foreign authorities · Difficult to maintain harmonization over time given changing technology
Information sharing and cooperation	· Builds on and institutionalizes lab self-regulation · Helps other jurisdictions get up to speed on regulation · Low cost and non-invasive	· Risks leaking dangerous or valuable information · Relatively toothless unless supplemented by other means such as fines
Modeling of rules and practices	· Shapes domestic frontier AI regulation of foreign states · Promotes UK leadership and UK sets the standard for frontier regulation	· Relies on foreign states to copy and implement rules effectively · Diffusion may fail because of internal obstacles in other states

The convening also found that the question of how to regulate frontier AI is structured by a set of apparent tradeoffs. Deciding how to manage these tradeoffs and determining whether there are ways to achieve or balance both values apparently being traded off will be key parts of crafting the bill. The main tradeoffs the convening identified are as follows:

1. **Specificity and flexibility in rules**: Because frontier AI models are developing so rapidly, regulators must be able to incorporate the science of safety and evaluation as it develops. However, the more flexible and often-changing regulation is, the more difficult it will be for companies to understand how to comply and for other jurisdictions to stay harmonized with the United Kingdom. Especially if international governance relies on a system of distributed regulators with mutual recognition, it will be necessary to develop a structure for accommodating regulatory evolution in a clear and predictable manner.
2. **Reach and reliance in means of regulation**: As discussed above, states must work together to regulate frontier AI, or the field will not be sufficiently covered. However, states may still face threats that emerge from within foreign jurisdictions that do not have sufficient regulatory oversight of frontier companies within their borders. As such, establishing some kind of constrained legal basis for reaching out to cover those companies in cases of emergency may be necessary.
3. **Science-forward and rule-based**: The science of frontier AI is rapidly developing, and regulators should ensure that their coverage of these systems is flexible and responsive to the science. However, they also should not entirely cede the field, and establishing some basic rules now would likely allow for more effective future regulation.

The international regulatory landscape of frontier AI is currently in flux as institutions form and disappear and rules are created and removed. The United Kingdom has an opportunity to shape frontier AI around the world through a well-crafted and robust bill establishing a regulatory framework, taking advantage of its accumulated expertise and strong position to secure the benefits of safety and promote innovation and economic opportunities. Because of its position of leadership in the field, the choices that the United Kingdom makes in creating its frontier AI bill will reverberate around the world. Thinking carefully about how the government can accomplish its goals and choosing the right mix of international regulatory tools, conscious of both their strengths and potential weaknesses, is an essential step forward in making sure that frontier AI is used for the good.

Table of contents

I. Introduction

Peter Kyle, UK Secretary of State for Science, Innovation and Technology, recently pledged that the government will put forward a bill aimed at regulating frontier AI during the current parliament.[1] If passed, that law would be the first in the world directly targeted at these cutting-edge models and systems that present both the greatest opportunities and

[1] Anna Gross & Stephanie Stacey, *UK will legislate against AI risks in next year, pledges Kyle*, FINANCIAL TIMES (Nov. 6, 2024), https://www.ft.com/content/79fedc1c-579d-4b23-8404-e4cb9e7bbae3 (quoting Peter Kyle, Secretary of State for Science, Innovation, and Technology).

the greatest risks for society.[2] This forthcoming UK frontier AI bill is already the focus of significant debate. However, the international dimensions of the bill have been relatively neglected despite their paramount importance. Any attempt to regulate frontier AI, which is being developed in companies scattered across jurisdictions[3] and used around the world, must confront questions about how to regulate internationally while preserving the benefits and opportunities that these models present.

To address this gap in the discussion, the Oxford Martin AI Governance Initiative convened a group of experts from industry, government, academia, and civil society to discuss how the UK could best shape its frontier AI bill to regulate internationally. The convening also sought to lay out some general lessons for international regulation that might be used by lawmakers in other jurisdictions. In particular, participants discussed the tools of international regulation that the UK could make use of in seeking to accomplish its goals of ensuring safety and promoting innovation, as well as how to structure the law to best make use of each of those tools.

Governments can seek to regulate outside their borders with varying degrees of directness, from full claims of authority over foreign entities to simply modeling good rules and practices in the hopes of encouraging others to adopt them in their own jurisdictions. Where international regulation is necessary, as it is in the case of frontier AI, developing a robust framework of international cooperation would best ensure safety and reduce the extent to which states have to try to regulate outside of their borders because each could rely on others. However, because such a cooperative framework does not yet exist and may not come into being in the near future, states must consider how much they need to develop tools for the regulation of foreign entities if serious risks from abroad come to light. Taking a position of leadership in the international order of frontier AI governance will enable the UK to best secure its own safety and reduce the extent to which it has to rely on unilateral forms of international regulation instead of cooperation, though the government should also consider whether it wants to maintain some unilateral tools as a backstop.

This report serves two main purposes: First, it seeks to clarify the choices of international regulation facing the UK government as it goes forward with its lawmaking process on the frontier AI bill. Second, it provides an abstract analysis of the kinds of decisions that must be made in the international regulation of frontier AI that may be useful to any authority

[2] Other laws, like the European Union AI Act and the Chinese regulations on Deep Synthesis Technologies and Generative AI take current frontier models into their ambits. However, these regulations were mostly or entirely designed and promulgated before the advent of frontier AI and are not specifically focused on these systems, though they include them. The United States has not passed a law regulating frontier AI but does have Executive Order 14110, which creates a set of governmental requirements around the technology. Furthermore, Executive Order 14110 may be revoked under the incoming Trump administration.

[3] The leading companies are currently concentrated in the United States and China, though other jurisdictions have companies that are also attempting to create leading models like Mistral in France.

seeking to regulate frontier AI. Participants in the convening found that the decision of how to regulate frontier AI is structured by a set of tradeoffs that lawmakers must navigate to best achieve safety and access the opportunities presented by frontier AI. For example, policymakers must determine how much to rely on others to regulate frontier AI versus how much to develop internal capabilities. They must also balance the promotion of clear, concrete rules that can be easily complied with by AI companies with the reality that the science of AI is developing and rules can quickly go out of date, making adaptability key. Determining how to navigate these tradeoffs and use the tools of international regulation to secure the benefits of AI is therefore a core task of lawmakers seeking to regulate frontier AI.

II. Background

The current UK government has said that it will put forward a bill regulating frontier AI. The exact shape of the bill is at this point still uncertain, but the government has suggested that it is interested in building a regulatory regime focused on frontier AI. It is unclear which body would be responsible for implementing the bill, but it is likely that existing sectoral regulators will deal with applications of frontier models that fall under their purview and with narrow AI systems built for their sectors. Sectoral regulators might also be granted new regulatory tools to help them respond to the advent of these new technologies in the law. The government has further indicated that as part of the new bill, it will make the UK AI Safety Institute (AISI) into an arm's-length body (ALB) and put the voluntary commitments made by AI companies at the AI Safety Summits at Bletchley Park and Seoul on a statutory footing. Together, these steps help concretize the progress in governance that has occurred in the past few years. Furthermore, the government has said that it wants to ensure that the benefits of AI, particularly innovation and economic growth, are not stifled by any law that it puts into place. The forthcoming bill will seek to balance these various objectives.

Outside the UK, the international landscape of AI governance is still shifting and taking shape. The EU AI Act General Purpose AI Code of Practice was recently released, providing more information about how that law will shape the regulation of AI.[4] The United States Executive Order 14110 on frontier AI will likely be revoked by the incoming Trump Administration, removing various requirements and also leaving the US AISI on

[4] *First Draft of the General-Purpose AI Code of Practice published, written by independent experts*, EUROPEAN COMMISSION (Nov. 14, 2024), https://digital-strategy.ec.europa.eu/en/library/first-draft-general-purpose-ai-code-practice-published-written-independent-experts.

shaky ground.[5] China has various regulations relevant to AI, but the general AI law draft circulated last year by the Chinese Academy of Social Sciences[6] still seems not to have been made into anything binding. International summits and other meetings through the United Nations, G7, OECD, and the like continue, but it is unclear exactly what the concrete outputs of those processes will be. As such, there is a significant opportunity for the UK to have a strong international influence through the passage of a frontier AI bill that creates a regulatory framework specifically for these technologies. Such a bill could shape the international regulatory environment both by directly regulating frontier AI models and indirectly by setting a standard for how regulation of these advancing technologies could be done.

III. Means of international regulation

There are six core ways that the UK can seek to regulate frontier AI internationally. These forms of regulation (extraterritorial application of laws, a de facto 'London Effect,' multilateral agreements, international harmonization, information sharing and cooperation with companies, and modeling of best practices) are discussed in depth below. Each means of regulation has its own advantages and disadvantages. Deciding which means or combination of means to use in which context is a necessary part of developing a regulatory framework for frontier AI. The UK is particularly well-positioned to regulate frontier AI given its early leadership in the field, its concentration of technical expertise (especially in the AISI), and its economic significance. Because of this position, any law that the UK produces will affect the international regulation of frontier AI, whether directly or indirectly. As such, the government should consider how substantive elements of the forthcoming bill can make use of the tools of international regulation laid out below to best accomplish the goals of securing safety while promoting innovation and growth.

Direct international regulation is likely to be a necessary part of developing a framework for governing frontier AI, but choices about how to structure domestic laws and institutions also have international effects. As such, the government must think about how the choices that it makes in setting up its internal governance system will affect those abroad as well. For example, the AISI is a government office that so far has mostly operated as a research authority advancing the science of AI governance, among other

[5] Madison Adler, *Trump likely to scale back AI policy with repeal of Biden order*, FEDSCOOP (Nov. 14, 2024), https://fedscoop.com/trump-likely-to-scale-back-ai-policy-with-biden-order-repeal/.
[6] *See* Kwan Yee Ng et al., *Translation: Artificial Intelligence Law, Model Law v. 1.0 (Expert Suggestion Draft) – Aug. 2023*, DIGICHINA (Aug. 23, 2023), https://digichina.stanford.edu/work/translation-artificial-intelligence-law-model-law-v-1-0-expert-suggestion-draft-aug-2023/.

things by cooperating with frontier AI companies to help them develop and perform evaluations of their models. The AISI has had significant influence abroad through its research and partnership work. The government has indicated that it will turn the UK AISI into an ALB in the forthcoming bill, increasing its institutional independence. AISI could then continue to advance the science of evaluations and engage in coordination and standard-setting, allowing the government to shape international AI regulation indirectly. Paired with a body to oversee frontier AI regulation, which could be set up under the forthcoming frontier AI bill, this approach would allow both direct and indirect regulation of frontier AI models through domestic law.

The six means of international regulation discussed during the convening can broadly be arranged on a scale from ‘harder,’ or more invasive, to ‘softer,’ or more relaxed. Some tradeoffs must be considered when choosing where to operate along this scale. Harder forms of regulation, including direct claims of jurisdiction over foreign entities, would, in most situations, give the regulating authority more control over the content of the regulation and likely over the foreign entities as well because they would force the regulated party to comply with rules written by the regulator. For example, a law that required that any entity anywhere in the world that produced a frontier AI model submit the model to evaluation by a UK evaluator would allow the UK to control the kind and quality of evaluations. However, such broad claims of extraterritorial authority are likely to be less effective in practice than in theory because of enforcement problems. Limiting claims of extraterritoriality to foreign AI providers that interact with or have significant effects on the UK might allow for the UK to strike a better balance.

On the other hand, a softer form of regulation, such as simply offering evaluations to companies that wanted to certify their safety to use as part of their marketing or for insuring their products, could leverage the high degree of capabilities within the UK AISI to achieve safety through cooperation. However, if companies refused to submit models for evaluation and there was no way to force them to do so, then the government would have little it could do about such refusals, which may create significant risks. Similarly, a soft approach, like modeling a regulatory regime in the hopes that other jurisdictions would follow suit in implementing such a regime themselves, would use leadership and persuasion as a mechanism for change. However, an approach like modeling would force the UK to rely on foreign jurisdictions’ understanding and adopting a regulatory regime in line with what the UK was modeling. Additionally, this kind of softer approach would depend on the foreign jurisdictions having sufficient technical capabilities to implement the regulatory regime themselves, something that may be difficult in light of the scientific challenges of AI regulation. The best international regulatory framework probably includes a combination of both harder and softer means of regulation that uses the strengths of the UK and other jurisdictions with high regulatory capacity while respecting and helping to develop the international regulatory capabilities of others.

The UK has gotten far through cooperation with frontier companies and is working to improve collaboration with foreign governance authorities. Building on that success is essential. However, the government should not shy away from seeking to develop a more robust form of international regulation of frontier AI. A strong frontier AI regulation that sets the tune of international governance of these systems will better promote safety and also provide a more stable regulatory environment for frontier AI companies, reducing the extent to which they have to comply with a patchwork of rules across jurisdictions and limiting uncertainty about where law will go next. Significant international frontier AI regulation would also reduce the risk that the UK or other states might face regulatory arbitrage by limiting the extent to which companies can seek to avoid being regulated by simply moving to other jurisdictions. Strong leadership on international regulation will give the UK long-lasting influence on how AI will be developed around the world.

a. Extraterritorial application of law

The first means of international regulation of AI is by direct extraterritorial application of UK law to foreign frontier AI developers. In this mode of regulation, rules created by domestic authorities would apply to foreign entities, and if the foreign entities violated those rules, then they would be subject to whatever means of enforcement the domestic authorities had at their disposal. The extraterritorial application of law is generally allowed only in exceptional circumstances as it is viewed as a violation of the sovereignty of others,[7] and states usually only regulate foreign entities with respect to significant contacts that they have with the state's jurisdiction. For example, corporations that establish residence or physical presence in the regulating jurisdiction or that continually avail themselves of the market of the jurisdiction are usually subject to that jurisdiction's laws, though they may be headquartered and mainly operate elsewhere. However, as the conduct of entities abroad increasingly spills across borders, some states, including the UK, have taken steps in some areas to attach jurisdiction to foreign entities that engage in conduct that has 'substantial and foreseeable harmful effects' (or similar language) within its borders.[8] Such an effects-based extension of jurisdiction remains constrained while also allowing the government to prevent significant harms that are not being dealt with by foreign authorities.

Extraterritorial jurisdiction raises the difficult question of extraterritorial enforcement. Because the entity being regulated exists outside the borders (and normal legal reach) of the regulating jurisdiction, the normal tools of law are more difficult to apply. Whereas a domestic company that refuses to comply with the law or to show up in court can be punished for such acts by seizing property held in the jurisdiction or excluding them from

[7] Menno T. Kamminga, *Extraterritoriality*, in MAX PLANCK ENCYCLOPEDIAS OF INTERNATIONAL LAW (2020).

[8] *Id.* See also Jonathan Ford, *UK antitrust reforms: a reversal of the UK's historical resistance to extraterritorial application post-Brexit*, LINKLATERS (May 13, 2022), https://www.linklaters.com/en/insights/blogs/linkingcompetition/2022/may/uk-antitrust-reforms.

the market, among other things, a foreign entity that has no real connection to the regulating jurisdiction cannot effectively be forced to conform with its law. This situation is generally a good one for the rule of law and respect for the sovereign prerogatives of states, but where harms can traverse borders easily, it presents some risks. As such, any claim of extraterritorial jurisdiction should be limited as much as possible only to situations in which the rules can be enforced, and the government must work to create cooperation with other jurisdictions around the world to get them to regulate potential sources of harm in their territories.

In the context of frontier AI regulation, the extraterritorial assertion of authority by the government may be necessary to reach foreign AI developers that present risks and that are not covered by any responsible foreign authority, but it should be limited to exceptional cases and then ideally only in cooperation with whatever foreign authorities do exist. For example, the extraterritorial assertion of UK authority might be necessary in the case of a model that can be used to create bioweapons or otherwise enable serious harm within the UK but that operates from a jurisdiction without a competent AI regulatory authority. In that kind of situation, a version of limited (for example, to "substantial and foreseeable effects" as discussed above) extraterritorial regulation would likely give the UK greater ability to reach out and regulate where necessary while still acting under the constraint of the law. Furthermore, establishing a real but limited domestic basis for extraterritorial regulation in the form of an effects-style provision could smooth the process of reaching out beyond the UK's borders and reduce concerns from foreign actors that the government is likely to abuse its powers or overreach in how it regulates abroad by cabining its authority to a small set of cases. The UK has already promulgated some effects-style extraterritorial legislation, for example in its competition laws, suggesting that this kind of limited claim of authority is acceptable where a need is shown.

Of course, even such a limited claim of extraterritorial authority would likely run into obstacles in certain cases, especially where other states see important interests implicated in the objects of regulation. China would be unlikely to respect the UK's claim to be able to regulate a Chinese frontier company that it saw as essential for its national security, for example. Similarly, where a state saw a frontier company as a national champion or where it had positioned itself as a kind of 'haven' for model providers to operate subject to less regulation, claims of extraterritorial jurisdiction by the UK or other foreign jurisdictions might be opposed by the government of the relevant state. These circumstances limit the potential of extraterritorial application alone to effectively regulate frontier AI and indicate that softer, cooperative forms of regulation are also needed to ensure safety, though some version of extraterritoriality might be necessary as a backstop.

As such, if the government chooses to create a regulatory body for frontier AI, it should consider giving that frontier regulator limited extraterritorial power via an effects provision in the bill. Combining the cooperative work of AISI and other authorities that coordinate with foreign evaluators and regulatory bodies with a UK regulator that could ensure that companies are following the rules and sticking to their commitments would allow the government to get the benefits of both soft and harder forms of international regulation.

b. A de facto London Effect

The UK could also seek to ensure the compliance of frontier AI models with UK law by establishing something like a de facto London Effect, whereby foreign model producers are incentivized to comply with domestic rules even while abroad. Previous examples from other jurisdictions include the California and Brussels effects.[9] This form of international regulation would be softer than direct extraterritorial regulation and thus less likely to create resistance but would also allow the UK to have more control over the content of regulation than international harmonization, modeling, or similar approaches that rely on other states regulating effectively.

In a London Effect, the government would create a set of rules that offer punishments or rewards for complying with domestic law, from fines to market exclusion to positive incentives like subsidies. For example, frontier AI companies that want to provide their models in the UK might be required to have their models undergo a variety of evaluations that prove the safety of their model before they are allowed to sell their model in the country. If they did not submit to the evaluations, they would be prevented from providing their AI in the UK or subject to fines until such compliance is achieved. Alternatively, demonstrating compliance could operate as a kind of partial shield against tort liability, creating a presumption of no-fault that would have to be overcome by a litigant in a suit. Depending on how the evaluations or other requirements affected the model providers, it might be more economically efficient for them to simply comply with the rules outside of their jurisdiction as well as inside it rather than "forking" their product to provide different versions for different rules, creating something like the California or Brussels Effects. So, if many of the regulations put forward by the UK affect pretraining or earlier stages of model development, where the costs of forking are untenably high, then the regulations are more likely to have a London Effect.

The extent to which a London Effect emerges will likely be governed by a variety of factors, some more and some less under the control of the UK.[10] One key area that the government will be able to control is the definiteness or specificity of the regulations that ground the

[9] See DAVID VOGEL, TRADING UP: CONSUMER AND ENVIRONMENTAL REGULATION IN A GLOBAL ECONOMY (1995) (describing the California Effect); ANU BRADFORD, THE BRUSSELS EFFECT: HOW THE EUROPEAN UNION RULES THE WORLD (2020) (describing the Brussels Effect).

[10] For a helpful list of factors that might increase or limit the extent to which a London Effect occurs, see Charlotte Siegmann & Markus Anderljung, *The Brussels Effect and Artificial Intelligence: How EU regulation will impact the global AI market*, CENTRE FOR THE GOVERNANCE OF AI 18 (Aug. 2022), https://cdn.governance.ai/Brussels_Effect_GovAI.pdf.

Effect. It seems likely that more specific regulations would generate more of an Effect because they would provide clearer compliance targets for AI companies to meet where general or flexible rules may not be as easy for companies to conform to, reducing the likelihood that they will comply beyond what is necessary. The question of specificity raises the potential for the dilemma discussed above between more specific regulations and more flexible and adaptable ones. It is possible that more adaptable regulations, which might be necessary to keep up with the advances in AI, are less capable of creating a London Effect because how to comply with them is less predictable over time. On the other hand, more specific regulations might end up no longer applying meaningfully to advancing technology, creating uncertainty from regulatory misfit. If the government chooses to pursue a London Effect, it should consider how to balance this specificity and flexibility as two important but potentially competing requirements of effective international frontier AI governance.

c. Multilateral agreements

The UK government could also use the forthcoming bill to seek to establish binding multilateral agreements with other states on frontier AI regulation, building on its existing leadership in international forums. Such binding agreements would provide relatively strong assurance to the UK that foreign states were effectively regulating their frontier AI models (especially if they included verification provisions that allowed each state to monitor the compliance of companies in other states through those other states' authorities) but may take a long time to come into being and would be subject to the compromises often necessary to establish such agreements. In principle, binding agreements among as many states as possible an optimal approach to international governance, ensuring that all views are represented and that every regulator is bought in while also achieving the force of binding law. However, the more participants in such a process, the more difficult and lengthier the process might be. The minimum required set of participants in such an agreement is those jurisdictions that play host to frontier AI companies, but overcoming conflicts between leading players like the United States and China might be too difficult over the next few years for a full binding treaty to emerge. There is some hope that existing processes in multilateral forums like the United Nations, OECD, and Safety Summits may continue to work toward something like binding agreements for AI safety. However, those processes may stall or go in different directions, particularly in light of the changing geopolitical situation.

However, this gap in existing multilateral agreements may present a useful opportunity for the government. If the UK establishes the initial framework for a multilateral frontier AI treaty or the standard for bilateral agreements on AI safety, such groundwork will likely have a lasting influence as it gains acceptance and support and other agreements are patterned off of it. The government should expand the UK's leadership in international AI diplomacy and seek to use that leadership to institutionalize the progress that has been made so far. Actions like making frontier AI companies' voluntary commitments (made at international forums at Bletchley Park and Seoul) into binding statutory commitments are good steps in this direction, but leading in existing multilateral forums and seeking to promote treaties involving smaller groups of states would also be a beneficial way to exercise the UK's particular position and talents.

d. International cooperation and harmonization

Beyond full multilateral agreements on frontier AI regulation among states, the UK should continue to pursue cooperation and harmonization with foreign authorities on evaluations, regulatory reciprocity, harmonization of standards, and similar fronts. Such work does not require the full inter-state agreement that an international treaty does but instead can be based on a simpler and, thus, easier-to-achieve form of cooperation among experts. Cooperation between the UK AISI and the US AISI or European AI Office, for example, can and does proceed on the basis of a memorandum of understanding or administrative arrangement rather than requiring the full treaty process, reducing the difficulties inherent in institutionalizing international cooperation.[11] Particularly where the regulatory regime must be expert and flexible, as is likely the case for frontier AI regulation, such a relatively low-level form of international cooperation might have significant advantages of adaptability over the higher-level international agreements discussed above. The science of AI is changing rapidly, and international regimes that were effective in one technological paradigm might look outdated in the next. As such, ensuring that regulators can work effectively together across borders while considering updates in the science of evaluation and AI more generally is necessary, and expert-level coordination and harmonization is one way to achieve that.

Significant international cooperation among expert groups like the AISIs already exists, and the government should explicitly seek to build on and further develop that cooperation. Directing the UK AISI to continue to proactively build collaboration with foreign AI regulators and evaluators would be a substantial step in the right direction, as would making it an ALB and increasing its independence. The UK AISI will likely be one of the most important ways that the UK government is able to influence international regulation of frontier AI, and it should be given the tools to effectively pursue that influence.

International harmonization is not a panacea to the problems of frontier AI regulation, however, and one key problem facing harmonization is how to maintain mutually acceptable standards, especially as the technology changes. Technological progress will mean that certain definitions and regulations developed for one paradigm of frontier AI

[11] *See, e.g.*, *Collaboration on the safety of AI: UK-US memorandum of understanding*, UK DEPARTMENT OF SCIENCE, INNOVATION, AND TECHNOLOGY (Apr. 2, 2024), https://www.gov.uk/government/publications/collaboration-on-the-safety-of-ai-uk-us-memorandum-of-understanding.

do not fit as well in the new paradigm, and governance authorities will have to consider how to respond to those shifts as they arise. While adapting the rules put forward by one authority to a new technology presents an already-difficult set of challenges, that adaptation will have to happen in each authority around the world in a way that allows the others to be sure that they are continuing to ensure the safety of frontier AI. The government should consider how to build a robust framework to support such harmonization across expert bodies, and taking the lead in the initial formation of that framework would give the UK significant influence over AI governance in the future.

e. Information sharing and cooperation

The government should also consider using the bill to more concretely direct UK evaluators and scientific authorities to share certain kinds of information with foreign regulators and other entities, including companies working on frontier AI, and build cooperation with them. The UK has a particularly strong concentration of talent in its AISI and has pioneered the science of AI safety. Sharing information about what it has learned would help other jurisdictions get up to speed and make sure that frontier AI companies covered by their law are developing and deploying their models safely by avoiding gaps in the coverage of effective regulatory regimes. Cooperating with foreign regulators, especially on the basis of a two-way transfer of information, would help the UK guide the science of safety elsewhere and shape what kinds of evaluations and regulations are occurring abroad.

UK authorities working with foreign companies to provide evaluations and other services would benefit those companies by giving them access to top-quality evaluations of the risks of their models while also promoting safety and encouraging those companies to develop and deploy their AI systems in the UK. Making the UK an attractive place for AI businesses and creating relationships with those businesses will ensure that evaluators can stay on top of the evolving frontier. The government should ensure that these relationships do not develop into a form of regulatory capture, but if handled well, for example, by having an evaluator like AISI work with the companies while a separate regulator enforces rules, then the benefits of collaboration could be significant.

The UK government should particularly consider sharing information with AI authorities that are just starting up in other parts of the world, including the developing world. AI will have a transformative effect around the globe, and places that lack the concentration of AI talent that the UK enjoys will particularly benefit from information sharing that allows them to more quickly develop their AI regulatory capacities and create robust and effective legal frameworks that allow for AI to be deployed safely in their jurisdictions. Those frameworks will benefit the UK in turn by improving the coverage of frontier AI regulation around the world, reducing the risk of a harmful model developed and deployed in a foreign jurisdiction harming the UK.

Certain kinds of information produced by UK authorities, including the UK AISI and other groups, are likely to be too sensitive to share broadly but should still be shared in some part with relevant authorities in other jurisdictions. These kinds of information, including evaluations related to sensitive national security topics as well as trade secrets that are essential to the success of frontier AI companies, should be protected against widespread publication, including to some partners who might otherwise receive access to information shared by UK authorities. To reduce the risks of sharing sensitive information, the government should invest in developing some kind of information-protecting verification system that evaluators could use to check the evaluation work of other jurisdictions without having to learn the full details of their processes. Such an approach would preserve the secrecy necessary for national security and sensitive commercial matters while also allowing for collaboration on key issues. Together, information sharing done right should make the UK both a more effective leader and a better home for frontier AI companies.

The government has also indicated that it will put the voluntary commitments to safety made by leading companies at the Bletchley Park and Seoul Summits onto a statutory footing. Such a move, especially made in continued cooperation with the companies, demonstrates the effectiveness of coordination in creating real progress in the governance of frontier AI. Frontier AI companies are aware of the dangers that their models could create and should be encouraged to continue to innovate in safety and mitigations to reduce those dangers. The government, acting as a partner to the companies and as a backstop to enforce commitments that have been made, can encourage such pro-social innovation and limit the incentive for a company to defect from its voluntary commitments in exchange for technological progress.

f. Modeling of rules and practices

Finally, the government should seek to model effective regulatory and evaluation practices for other jurisdictions. Through this mechanism, the UK would seek to export its regulatory practices through their acceptance by other jurisdictions that have their own frontier AI governance rules and practices. Regulatory diffusion of this kind, whether simply through modeling *per se* or by advising regulators elsewhere on how to govern frontier AI, would allow the UK to expand its international reach while limiting the burden that it faces itself and reducing the need to directly seek to regulate outside its borders. Modeling good regulation, especially in a fast-moving and technical field such as frontier AI, may have significant effects elsewhere as authorities in other jurisdictions seek to quickly come up with ways to effectively govern AI and draw on existing patterns to do so.

However, modeling is unlikely to be a full guarantee of safety, and relying on it would likely limit the extent to which the UK could influence frontier AI abroad. Modeling requires other jurisdictions to understand and correctly copy or adapt the regulations put in place by the modeler. Where the development and application of those regulations require significant expertise, as it does in the context of frontier AI, there may be limits on how effectively other jurisdictions can shape their own regulations after the pattern of the modeler. Even if foreign jurisdictions were able to simply copy and paste the UK's laws into their own codes, without the talent and regulatory capacity needed to implement these laws effectively, they would not be able to fulfill their function. But it is unlikely that such direct copying would even be possible because foreign jurisdictions will have to go through their own lawmaking processes and likely have to compromise on domestic political considerations that make complete adoption of the model regulation difficult. The relatively limited modeling that has occurred between the European Union, United States, UK, and China with respect to other forms of technology regulation suggests that modeling may not occur to a significant extent in the case of frontier AI. Taken together, these conditions limit the effectiveness of modeling laws internationally and mean that modeling should be part of a toolkit of international regulation along with other more direct forms.

IV. Shaping the bill to achieve the goals of international regulation

The forthcoming frontier AI bill will have international implications because it will be one of the first key pieces of frontier AI regulation and because of the UK's position of leadership in the field. As such, it should be written so that the various means of international regulation discussed above are best used to achieve the government's goals of safety and growth. The convening focused on four key elements of the frontier AI bill that will have particularly significant international implications: the definition of frontier AI and how it can change, the legal coverage of the bill, how the bill deals with direct regulation and enforcement, and how the bill treats open-source providers and downstream users. Debates over the content of each of these provisions are likely to be contentious, and it is worth considering how choices about these provisions will affect the international landscape. Balancing adaptability and consistency, reach and reliance on others, and leaving space for innovation with using the law to provide a stable and predictable space for growth, will be key challenges in writing the bill, but if done effectively, the bill will shape the field of frontier AI into the future and help the UK secure its goals of safety and growth.

a. Technical definition of frontier AI

The bill should include a definition of frontier AI that allows for coordination across jurisdictions to make sure that the UK can rely on regulation of frontier AI that happens elsewhere. Avoiding gaps in coverage of frontier AI will require consistent governance across jurisdictions. Especially if the government decides not to claim direct authority over frontier AI companies outside of its borders or the usual reach of its jurisdiction, developing a coordinated framework for regulation with foreign model providers will be necessary. However, coordinating the definition of frontier AI across borders might be particularly difficult given that the science of AI is changing constantly, so definitions that fit one year might not fit the next. Every jurisdiction would then need to update its definition of frontier AI in a way that allows for harmonization to persist across time.

Defining frontier AI is a difficult challenge, and each of the proxies that have been proposed for covering the topic, including compute used, cost of model production, and model capabilities, will face obstacles.[12] Furthermore, changes in technology could lead to definition misfit over time. For example, a definition of frontier AI that relies entirely on compute resources used during training may become less effective if it turns out that progress at the frontier ends up being driven by algorithmic innovation or post-training compute (a real possibility presently),[13] inputs that are not captured by the compute-based frontier model definition. Similarly, a cost-based definition may be effective if the only dangerous models are those that are expensive. But if the cost of creating a dangerous model falls over time, as it should given the falling cost of inputs, then a trigger that relies on training costs may also fail. Capabilities-based definitions are least likely to fall prey to misfit but face a different problem, which is that they are difficult to use as a trigger for evaluation or regulation. If a model is only evaluated for dangerousness if it reaches some capabilities threshold, but the capabilities of a model cannot be determined until after it has been evaluated, then the definition will not work. Furthermore, simply covering general models would exclude frontier models that are specialized in a particular area, for example, something like Google DeepMind's AlphaFold that might be able to create harmful kinds of molecules. It may also be the case that the next generation of AI is more oriented around systems than particular models, such that covering models misses some sources of risk.

[12] The EU AI Act and US EO 14110 both use compute thresholds as the trigger for coverage under their respective governance regimes, though the AI Act used 10^{25} floating point operations (FLOPs) while EO 14110 uses 10^{26}. California's SB 1047, recently vetoed by Governor Gavin Newsom, used cost alongside compute as its threshold, covering models whose pretraining cost more than $100 million and those whose finetuning cost more than $10 million. Various Chinese regulations cover models that have certain capabilities, including generating images and the like, and the EU AI Act also refers in Art. 51 to "high-impact capabilities" in defining general purpose AI models with systemic risks.

[13] Recent reporting that scaling may have hit a wall suggests that algorithmic innovation and post-training compute will become increasingly important. *See* Rachel Metz, Shirin Ghaffary, Dina Bass, and Julia Love, *OpenAI, Google and Anthropic Are Struggling to Build More Advanced AI*, BLOOMBERG (Nov. 13, 2024), https://www.bloomberg.com/news/articles/2024-11-13/openai-google-and-anthropic-are-struggling-to-build-more-advanced-ai; Stephanie Palazzolo, Erin Woo, & Amir Efrati, *OpenAI Shifts Strategy as Rate of 'GPT' AI Improvements Slows*, THE INFORMATION (Nov. 9, 2024), https://www.theinformation.com/articles/openai-shifts-strategy-as-rate-of-gpt-ai-improvements-slows.

Furthermore, it is possible that focusing on AI models as the targets of regulation, rather than on AI systems or architectures, will cause the definition to miss the models that must be regulated for the government to achieve its goals. If it turns out that the frontier is pushed forward through the development of model architectures rather than increasing the inputs used in pretraining models, for example, then compute- and cost-based definitions of frontier AI will likely not cover the key models. Research by AISI and similar institutions involving cooperation with leading companies will help regulators keep tabs on where the most dangerous frontier AI models are, and adapting to respond to these shifting sources of risk is essential.

The best approach to a definition likely involves combining several different dimensions of frontier AI with the ability to change the definition over time in response to changing technology. The government must decide which authority should be given the responsibility of updating the definition of frontier AI over time and determine what process would be required to make a change in the definition. However, it is essential that, regardless of what structure is chosen to solve this problem, the ability to engage in international harmonization efforts be preserved. Communicating with other frontier regulators and ensuring that the definition of frontier AI used in the UK and abroad is consistent such that each jurisdiction can rely on regulation and evaluations done by the others is an essential criterion for effective international regulation of AI.

b. Legal coverage

The government should shape the bill to clearly specify when it applies to domestic and foreign entities. The bill's technical definition of frontier AI, whether in the bill or adopted by the relevant authority through a standard-setting mechanism set up through the bill, should be one threshold for application, but the government should also clarify the legal criteria for when a frontier AI company, model, or system is covered by the law. Deciding which foreign providers are covered by the bill and when a foreign entity becomes covered is essential to creating a clear and consistent regulatory framework. At a baseline, the bill should cover any foreign entity that avails itself of the UK market by providing models or other services in the UK. Such a baseline is consistent with most kinds of regulation and is a normal exercise of governmental power. The leading American frontier AI companies are all registered in the UK, as are some Chinese companies, and would straightforwardly be covered by this kind of application framework.

The government should also consider more expansive legal coverage than the default, for example, by including an 'effects' provision in the bill.[14] Because some frontier AI companies may not directly interact with the UK or another jurisdiction that has a regulator able to certify the safety of the models or systems produced by the company, it may be necessary to allow the UK regulator to reach outside the borders of the UK in

[14] See *supra* Section III Part a.

exceptional circumstances to cover such a company. An ‘effects’ provision, which would allow the regulator to cover companies that have a substantial and foreseeable effect within the UK even if they do not directly interact with it, would provide a legal basis for such coverage while also limiting it and avoiding a full claim of extraterritorial authority over frontier AI. Similar provisions have been used in areas like competition law in the UK and elsewhere to allow authorities to fulfill their missions when harm is caused by actors outside the country.[15] The language of the provision could be written such that the actual use of an effects provision would be constrained to circumstances in which substantial harm is particularly likely, and no partner authority can cover the harm, limiting the extent to which other jurisdictions see the bill as a violation of their sovereignty.

c. Regulation and enforcement

The government should consider the use of fines, market exclusion, and other tools to enforce rules and regulations made to govern frontier AI. However, the obstacles to creating and funding such a regulator and ensuring that it is staffed with sufficient talent are significant. The convening did not come to any conclusion about whether establishing a dedicated frontier regulator or upskilling existing sectoral regulators to deal with frontier AI would be best, nor whether AISI should be given its own regulatory capabilities. However, it is worth laying out some of the advantages for international regulation of a dedicated frontier AI authority separate from AISI to help think through how to address the problem of frontier AI regulation because the creation of a separate authority is most distinct from the existing framework.

Among the key advantages of a dedicated frontier AI regulator for international regulation as opposed to upskilling existing sectoral regulators are the concentration of talent and resources, the reduction of the number of entities that foreign regulators and companies have to interact with, and the facilitation of modeling of good practices. A dedicated frontier AI regulator would allow the UK to concentrate talent and resources in fewer bodies and avoid duplicating work and effort across the regulatory system while avoiding reducing the role of AISI as an independent expert. The UK currently benefits from high levels of talent and expertise in the frontier AI space, but the amount of available talent for frontier AI regulation is relatively limited. As such, creating a system in which individual sectoral regulators have to each build out frontier AI expertise would risk leaving them competing for talent and creating shortfalls in certain areas. Furthermore, because there are a limited number of frontier AI companies that are creating potentially dangerous products, having one authority that can regulate each of them in a way that creates guarantees that can be relied upon down the chain of regulation by sectoral authorities would allow for the different authorities to maximize comparative advantages and do what they are best at. Furthermore, a regulatory framework in which one top-level

[15] Ford, *supra* note 8.

frontier AI regulator could help assure the safety of frontier systems would allow users down the stack to incorporate frontier models into their products and systems without worrying as much about liability, promoting the adoption and deployment of safe systems while maintaining sectoral regulation.

A dedicated frontier regulator likely allows for better use of soft forms of international influence. A frontier AI regulator would be able to work with foreign frontier AI authorities by giving them a clear point of contact that they could work with in coordinating and harmonizing international regulatory regimes. Because so much of the work of regulating frontier AI will be done in international forums, having a clear regulatory voice that can speak alongside expert groups like AISI in such forums would help establish the leadership of the UK in the field. Foreign frontier AI companies would face reduced regulatory burdens if they only had to work with one authority rather than dealing with many diverse sectoral regulators. That reduced burden would make them more likely to invest in the UK and increase the extent to which other forms of soft influence would work to advance frontier AI regulation.

Again, whether the above advantages of a dedicated frontier regulator are sufficient to overcome its potential costs and downsides is a difficult question that the convening did not seek to resolve. This analysis is intended to lay out some of the ways in which choices about domestic regulatory regimes can have significant international influence and help lawmakers think through how to structure their regulatory regimes to best achieve their goals.

d. Open-source AI developers

Participants in the convening also raised the problem of international regulation of open-source model providers. One of the key benefits of the frontier AI regulatory scheme is that it puts the burdens of regulation on the leading frontier AI companies that are best equipped to handle them and allows for the concentration of regulatory resources on the places where harms are most likely to arise. However, open-source providers represent a particular difficulty for this kind of scheme as they may not be attached to any jurisdiction with a frontier AI regulatory regime and can release their models for use by anyone in the world (even in jurisdictions that otherwise have regulation). Thus, for example, if OpenAI's frontier model is provided in the UK, then the UK frontier AI authority can go to OpenAI and require that it perform various evaluations or mitigations or face fines or other kinds of punishment. But if a model is released by an open-source lab in a jurisdiction that does not have a regulator that cooperates with the UK authority and the model is accessible by anyone anywhere with an internet connection, then the model or its effects might reach the UK anyway, and the system of coordinated governance would have failed. Regulating these kinds of open-source AI developers would be particularly difficult if they are beyond the legal and practical reach of authorities, either because those authorities do not have the jurisdiction to reach into the foreign territory (and no way to regulate the frontier AI provider in their own territory) or because they cannot effectively use the tools of enforcement to affect the lab (for example because the open-source lab has no assets that can be reached by any relevant authority).

Of course, there are also significant benefits of open source that should be protected by the forthcoming bill. Open source has a long tradition of promoting technology for the public good, and open-source collaboration on AI governance would allow for more effective research into the science of safety and limit the risks that frontier AI is controlled by a few small companies that could use their positions to harm others. Thus, a bill that effectively seeks to outlaw open source would go too far and perhaps do more damage than good. Instead, the bill should seek to encourage open-source and downstream development and uses of AI where such developments are consistent with a minimum standard of safety.

One possible solution to the problem of how to promote open source while guaranteeing safety would be using AISI or another expert technical authority to provide free evaluations to open-source AI developers and help ensure safety through cooperation with them. Providing free access to such services, especially to open-source projects that are not undertaken for profit, would reduce the burden that such groups would otherwise face in guaranteeing the safety of their models and make open-source AI more possible by increasing its reliability and making it more attractive to downstream users. Furthermore, by building relationships between open-source developers and governmental authorities, this kind of offering would improve communication and allow regulators to advise developers and shape how they think about safety, even if not directly regulating them. International coordination through informal agreements, advising, and information sharing would be the best way to seek to square this circle and promote effective governance.

V. Conclusion

The UK government has an opportunity to shape international regulation of frontier AI through its forthcoming bill. International regulation is necessary given the global character of frontier AI and its effects, and designing the bill with international influence in mind will best allow the government to achieve its objectives of safety, innovation, and growth for the UK. The six means of international regulation laid out in this report represent different ways that the government could seek to influence the international governance regime that is developing in the field of frontier AI, as well as the companies and other non-state actors that are pushing the field forward. Selecting the right mix of tools to use and ensuring that the regulatory framework developed is flexible enough to adapt to changes in technology and broader context are necessary steps toward an effective governance regime. The choices made in designing the frontier AI bill will have international effects, both directly and indirectly, as the UK’s particularly strong international position and concentration of talent and expertise make it an influential player in the space. Effective international regulation will help the government achieve its goals and ensure that the benefits of frontier AI are unlocked while its potential downsides are limited.