Supplementary Information

Table 1: Assessment of measures' feasibility and compatibility with open-source weights

Category	#	Measure	Feasible?	Compatible with open-source weights?
Responsible Development	1	Expanded Dual-Use Review	Yes	Yes
	2	Model Licensing for Training or Release	Some challenges	Yes
	3	Expanded Developer Liability	Some challenges	Yes
	4	Voluntary Commitments	Yes	Yes
	5	Export Controls	Some challenges	Yes
	6	Publication Norms	Yes	Yes
Risk Assessment	**7**	**Model Evaluations for Dangerous Capabilities**	Yes	Yes
	8	Red Teaming	Yes	Yes
	9	Monitoring for Misuse	Some challenges	Requires Structured Access
Transparency	10	Impact Statements	Yes	Yes
	11	Information-Sharing with Regulators	Yes	Yes
	12	**Vulnerability Reporting**	Yes	Yes
	13	Watermarking	Some challenges	May be fine-tuned away
Access Management	14	Data Curation	Yes	Yes
	15	Data Use Agreements	Yes	Yes
	16	**Structured Access**	Some challenges	No
	17	**Know Your Customer**	Some challenges	Requires Structured Access
	18	**Nucleic Acid Synthesis Screening**	Yes	Yes
	19	Input/Output Filtering	Yes	May be fine-tuned away
Cybersecurity	20	Database Security	Some challenges	Yes

Category	#	Measure	Feasible?	Compatible with open-source weights?
	21	Securing Weights	Some challenges	No
	22	**Securing Lab Equipment**	Yes	Yes
Investing in Resilience	**23**	**Model-Sharing Infrastructure**	Yes	Yes
	24	Public Compute	Yes	Yes
	25	Fund Countermeasures	Yes	Yes

Discussion: The decentralized and non-commercial nature of BDT development makes it harder to implement and target regulation

Unlike frontier foundation models, a wide variety of BDTs are developed by a range of actors across the world, often in not particularly well-resourced academic or start-up labs. This decentralization makes it much more difficult to implement regulatory proposals that may be appropriate when dealing with only a few large technology companies. In particular,

- **It is harder to target regulation in a proportionate way**: the breadth of different BDTs means that it is not easy to delineate which tools pose the most risk. This increases the likelihood that regulation has the unintended effect of slowing down beneficial science. While this is also an issue for other types of AI, the lack of effective compute governance options makes this more concerning.
- **It is more difficult to track compliance**: with more actors involved, any hypothetical regulator would have to spend more time and money ensuring that developers are following the rules. This concern is compounded by the fact that a malicious actor will try and conceal evidence of misuse. In contrast, with only a few companies capable of developing a foundation model at the level of GPT-4, it is straightforward for regulators to identify those companies whose models might pose risks.
- **There is a greater risk of regulatory arbitrage**: with BDT development spread across the world, many different countries may need to harmonize their legislation to effectively mitigate risks. This is also true for non-biology AI, to an extent, but the overwhelming concentration of frontier foundation model development in the US has meant that it is less of an immediate concern.
- **It is harder to consult all model developers to develop best practice**: in contrast with the White House Voluntary Commitments, which initially applied to only seven advanced AI companies (White House, 2023), developing best practices for BDT development is likely to be slower and more difficult. Academic conferences and partnerships like RosettaCommons will be key to developing such best practices.

Moreover, that many BDT developers lack the resources of large technology companies means that they cannot as easily implement potentially crucial cybersecurity features, both through lack of funding and expertise. We explicitly provide cybersecurity recommendations in Section 4.5, several of which will likely require developers and governments to work closely together.