Regulation Priorities for Artificial Intelligence Foundation Models

Matthew R. Gaske[*]

ABSTRACT

This Article responds to the call in technology law literature for high-level frameworks to guide regulation of the development and use of Artificial Intelligence (AI) technologies. Accordingly, it adapts a generalized form of the fintech Innovation Trilemma framework to argue that a regulatory scheme can prioritize only two of three aims when considering AI oversight: (1) promoting innovation, (2) mitigating systemic risk, and (3) providing clear regulatory requirements. Specifically, this Article expressly connects legal scholarship to research in other fields focusing on foundation model AI systems and explores this kind of system's implications for regulation priorities from the geopolitical and commercial competitive contexts. These models are so-named because they have a novel ability to easily apply their resources across a broad variety of use cases, unlike prior AI technologies. These systems, such as OpenAI's ChatGPT or Alphabet's LaMDA, have recently rocketed to popularity and have the potential to fundamentally change many areas of life. Yet, legal scholarship examining AI has insufficiently recognized the role of international and corporate competition in such a transformational field. Considering that competitive context and the Trilemma, this Article argues from a descriptive perspective that solely one policy prioritization choice is needed: whether to emphasize systemic risk mitigation or clear requirements, given that prioritizing innovation is effectively a given for many governmental and private actors. Next, regulation should

[*] J.D., Vanderbilt University Law School, 2016; M.B.A., University of Michigan Ross School of Business, 2023. The views expressed in this Article are the author's and are no indication of the views of any employers or institutions to which the author has, had, or will have an affiliation. While the author works or has worked with an early-stage venture capital firm focusing on artificial intelligence and similar subject matter, he neither has nor had a financial interest in that fund or its investments. © 2023, Matthew R. Gaske. Thank you is particularly warranted to Yesha Yadav, Daniel Sokol, Brian Wu, Cheng Gao, and the VANDERBILT JOURNAL OF ENTERTAINMENT AND TECHNOLOGY editorial team.

prioritize systemic risk over clarity because foundation models present a substantive change in the potential for, and nature of, systemic disruption. Finally, the Article considers ways to mitigate regulators' lack of legal clarity. It argues instead, in light of the Trilemma's application, for use of a sliding scale of harm-based liability for AI providers when reasonably implementable, known technological advances could have prevented injury. This tradeoff thus promotes innovation and mitigates systemic risk from foundation AI models.

TABLE OF CONTENTS

I. INTRODUCTION

The utility of advanced artificial intelligence systems has seized public attention, from the introduction of natural-language-prompted picture generation with OpenAI's DALL-E to Microsoft's incorporation of the ChatGPT system into its Excel and Bing products.[1] The recently developed interactive AI systems like ChatGPT are examples of Large Language Models (LLMs), which are "deep learning algorithm[s] that can recognize, summarize, translate, predict and generate text and other content based on knowledge gained from massive datasets."[2] Indeed, these models are not just limited to displaying words on a screen. Rather, they are examples of AI's utility as "a general-purpose technology" that can be repurposed for specialized tasks, such as allowing a user to prompt a robot with directions to manipulate its environment.[3] Concordantly, developers can build "general agent" AI to accomplish a broad variety of physical, visual, or written tasks.[4]

[1] *See DALL·E: Creating Images From Text*, OPENAI (Jan. 5, 2021), https://openai.com/research/dall-e [perma.cc/N2E9-NLDB] (last visited Mar. 30, 2023); Jonathan Vanian, *Microsoft Adds OpenAI Technology to Word and Excel*, CNBC (Mar. 16, 2023, 2:23 PM), https://www.cnbc.com/2023/03/16/microsoft-to-improve-office-365-with-chatgpt-like-generative-ai-tech-.html [https://perma.cc/4KWD-5476].

[2] Angie Lee, *What Are Large Language Models Used For?*, NVIDIA (Jan. 26, 2023), https://blogs.nvidia.com/blog/2023/01/26/what-are-large-language-models-used-for/ [https://perma.cc/8A4P-P7Q6].

[3] Nicholas Crafts, *Artificial Intelligence as a General-Purpose Technology: an Historical Perspective*, 37 OXFORD REV. ECON. POL'Y 521, 521 (2021) (defining "general-purpose technology . . . as a single generic technology, recognizable as such over its whole lifetime, that initially has much scope for improvement and eventually comes to be widely used, to have many uses, and to have many spillover effects." (internal quotation marks omitted)); Erik Brynjolfsson, Daniel Rock & Chad Syverson, *Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics*, 19–20 (Nat'l Bureau of Econ. Rsch., Working Paper No. 24001, 2017), https://www.nber.org/system/files/working_papers/w24001/w24001.pdf [https://perma.cc/9GXC-SVXN] (© 2017 by Erik Brynjolfsson, Daniel Rock, and Chad Syverson. All rights reserved.) (arguing that artificial intelligence is more properly understood and contextualized as a "general purpose technolog[y]" and echoing that classification's traits of "be[ing] pervasive, able to be improved upon over time, and be able to spawn complementary innovations"); Danny Driess, Fei Xia, Mehdi S. M. Sajjadi, Corey Lynch, Aakanksha Chowdhery, Brian Ichter, Ayzaan Wahid, Jonathan Tompson, Quan Vuong, Tianhe Yu, Wenlong Huang, Yevgen Chebotar, Pierre Sermanet, Daniel Duckworth, Sergey Levine, Vincent Vanhoucke, Karol Hausman, Marc Toussaint, Klaus Greff, Andy Zeng, Igor Mordatch & Pete Florence, *PaLM-E: An Embodied Multimodal Language Model*, 202 PROC. MACH. LEARNING RSCH. 8469–88 (2023) https://palm-e.github.io/ [https://perma.cc/2P4X-E44H].

[4] *See* Scott Reed, Konrad Zolna, Emilio Parisotto, Sergio Gomez Colmenarejo, Alexander Novikov, Gabriel Barth-Maron, Mai Gimenez, Yury Sulsky, Jackie Kay, Jost Tobias Springenberg, Tom Eccles, Jake Bruce, Ali Razavi, Ashley Edwards, Nicolas Heess, Yutian Chen, Raia Hadsell, Oriol Vinyals, Mahyar Bordbar & Nando de Freitas, *A Generalist Agent*, TRANSACTIONS MACH. LEARNING RSCH. 2–3 (2022), https://arxiv.org/abs/2205.06175 [https://perma.cc/8L9T-E4VB]. The furthest reach of this technology has been described as "a

First, a note on definitions and scope. There is no definition of "artificial intelligence" that is both precise and universally accepted.[5] For illustration, one example "is the hypothetical ability of a computer to match or exceed a human's performance in tasks requiring cognitive abilities, such as perception, language understanding and synthesis, reasoning, creativity, and emotion."[6] This Article does not seek to craft a perfect definition. Instead, terms such as "AI systems" or similar verbiage here merely name the relationship between (1) a software model that updates predictions as it acquires information to produce an output, and (2) that iterative process's connection to actions either another software system or some external entity completes.[7]

AI tools are rapidly becoming ubiquitous across many facets of modern life because of the recent surge in computing power,[8] abundance of available data,[9] and hardware innovations like graphics processing units.[10] These developments have the potential to connect users to "the wisdom of crowds, the power of information technology, and the

significant step towards" thinking like a human. Sébastien Bubeck, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuanzhi Li, Scott Lundberg, Harsha Nori, Hamid Palangi, Marco Tulio & Ribeiro Yi Zhang, *Sparks of Artificial General Intelligence: Early Experiments with GPT-4*, MICROSOFT RSCH. 4 (2023), https://arxiv.org/pdf/2303.12712.pdf [https://perma.cc/KY43-FVDM].

[5] *See, e.g.*, Michael Guihot, Anne F. Matthew & Nicolas P. Suzor, *Nudging Robots: Innovative Solutions to Regulate Artificial Intelligence*, 20 VAND. J. ENT. & TECH. L. 385, 393–96 (2017) (considering various definitions); William Magnuson, *Artificial Financial Intelligence*, 10 HARV. BUS. L. REV. 337, 342–345 (2020) (describing definitional difficulties and suggesting a "spectrum").

[6] Paul Grimm, Maura R. Grossman & Gordon V. Cormack, *Artificial Intelligence as Evidence*, 19 NW. J. TECH. & INTELL. PROP. 9, 14 (2021).

[7] For an instructive discussion of the difficulties in defining a novel technology for legal purposes, using the "robot" context, *see* Bryan Casey & Mark Lemley, *You Might Be A Robot*, 105 CORNELL L. REV. 287, 324–40 (2020) ("Definitions can fail: (1) when drafting laws from scratch to cover robots; (2) when robots interact with existing laws drafted with either people or traditional machines in mind; and (3) when robots defy the definitional bounds of existing regulatory bodies." Also noting issues with "overbroad" definitions).

[8] Jaime Sevilla, Pablo Villalobos, Juan Felipe Cerón, Matthew Burtell, Lennart Heim, Amogh B. Nanjajjar, Anson Ho, Tamay Besiroglu & Marius Hobbhahn, *Compute Trends Across Three Eras of Machine Learning,* ARXIV 3–5 (Mar. 9, 2022), https://arxiv.org/pdf/2202.05924.pdf [https://perma.cc/VBY5-2UMQ] (providing graphs demonstrating computational intensity of a "Pre Deep Learning Era," a "Deep Learning Era," and a "Large-Scale Era."). Notably, this Article focuses on what is often called "Narrow AI," though this description has been criticized "because it rests on dissimilar considerations of breadth and strength." Guihot et al., *supra* note 5, at 393, 397 (proposing alternative, "risk"-based classification approach).

[9] Andreas Kaplan & Michael Haenlein, *Rulers of the World, Unite! The Challenges and Opportunities of Artificial Intelligence*, 63 BUS. HORIZONS 37, 40 (2020).

[10] Nicole Kobie, *NVIDIA and the Battle for the Future of AI Chips*, WIRED (June 17, 2021¸ 10:10 AM), https://www.wired.co.uk/article/nvidia-ai-chips [https://perma.cc/9QYA-FU37]; *see also* Grimm et al., *supra* note 6, at 19.

precision of scientific methods into an iterative learning process" that accommodates individual preferences.[11]

Society has become more interconnected through digitalization and the related automation of digital tasks.[12] Yet because AI systems' capabilities have expanded dramatically, it is reasonable to consider AI systemic risk because of its status as an "enabler" technology that is, or can become foundational to, the mechanics of people's everyday lives.[13] In fact, researchers have recognized AI's capability leap with systems such as LLMs, using the term "foundation model" to denote an AI "model that is trained on broad data (generally using self-supervision at scale) that can be adapted . . . to a wide range of downstream tasks."[14]

[11] Rory Van Loo, *Digital Market Perfection*, 117 MICH. L. REV. 815, 830 (2019).

[12] *See* Mark Knell, *The Digital Revolution and Digitalized Network Society*, 2 REV. EVOLUTIONARY POL. ECON. 9, 11 (2021). For relevant distinctions between autonomy and automation, consider Arnault Ioualalen & Baptiste Aelbrecht, *The Autonomy of AI, Can We Really Talk About It?*, NUMALIS (Nov. 6, 2020), https://numalis.com/publications-26-the_autonomy_of_ai_can_we_really_talk_about_it.php [https://perma.cc/3RF4-C9G8].

[13] HENRY KISSINGER, ERIC SCHMIDT & DANIEL HUTTENLOCHER, THE AGE OF AI AND OUR HUMAN FUTURE, 4 (2021); *see also* Kimberly Houser & Anjanette Raymond, *It Is Time to Move Beyond the 'AI Race' Narrative: Why Investment and International Cooperation Must Win the Day*, 18 NW. J. TECH. & INTELL. PROP. 129, 130 (2021) (describing the scope of "the Fourth Industrial Revolution"); Han-Wei Liu & Ching-Fu Lin, *Artificial Intelligence and Global Trade Governance: A Pluralist Agenda*, 61 HARV. INT'L L.J. 407, 408–09 (2020).

[14] RISHI BOMMASANI, DREW A. HUDSON, EHSAN ADELI, RUSS ALTMAN, SIMRAN ARORA, SYDNEY VON ARX, MICHAEL S. BERNSTEIN, JEANNETTE BOHG, ANTOINE BOSSELUT, EMMA BRUNSKILL, ERIK BRYNJOLFSSON, SHYAMAL BUCH, DALLAS CARD, ANNIE CHEN, KATHLEEN CREEL, JARED QUINCY DAVIS, MOUSSA DOUMBOUYA, ESIN DURMUS, STEFANO ERMON, RODRIGO CASTELLON, NILADRI CHATTERJI, DOROTTYA DEMSZKY, JOHN ETCHEMENDY, LI FEI-FEI, CHELSEA FINN, TREVOR GALE, LAUREN GILLESPIE, KARAN GOEL, CHRIS DONAHUE, KAWIN ETHAYARAJH, NOAH GOODMAN, SHELBY GROSSMAN, NEEL GUHA, DANIEL E. HO, JENNY HONG, DAN JURAFSKY, PRATYUSHA KALLURI, OMAR KHATTAB, ANANYA KUMAR, TATSUNORI HASHIMOTO, KYLE HSU, JING HUANG, SIDDHARTH KARAMCHETI, PETER HENDERSON, THOMAS ICARD, GEOFF KEELING, JOHN HEWITT, SAAHIL JAIN, FERESHTE KHANI, XIANG LISA LI, SUVIR MIRCHANDANI, TENGYU MA, ALI MALIK, ERIC MITCHELL, ZANELE MUNYIKWA, PANG WEI KOH, FAISAL LADHAK, XUECHEN LI, MARK KRASS, RANJAY KRISHNA, ROHITH KUDITIPUDI, MINA LEE, TONY LEE, JURE LESKOVEC, ISABELLE LEVENT, CHRISTOPHER D. MANNING, SURAJ NAIR, AVANIKA NARAYAN, BEN NEWMAN, ALLEN NIE, JUAN CARLOS NIEBLES, HAMED NILFOROSHAN, DEEPAK NARAYANAN, JULIAN NYARKO, GIRAY OGUT, LAUREL ORR, ISABEL PAPADIMITRIOU, JOON SUNG PARK, CHRIS PIECH, EVA PORTELANCE, CHRISTOPHER POTTS, ADITI RAGHUNATHAN, ROB REICH, FRIEDA RONG, YUSUF ROOHANI, CAMILO RUIZ, JACK RYAN, CHRISTOPHER RÉ, SHIORI SAGAWA, KESHAV SANTHANAM, ANDY SHIH, KRISHNAN SRINIVASAN, HONGYU REN, DORSA SADIGH, ALEX TAMKIN, ROHAN TAORI, ARMIN W. THOMAS, FLORIAN TRAMÈR, ROSE E. WANG, WILLIAM WANG, BOHAN WU, JIAJUN WU, YUHUAI WU, SANG MICHAEL XIE, MICHIHIRO YASUNAGA, JIAXUAN YOU, MATEI ZAHARIA, MICHAEL ZHANG, TIANYI ZHANG, XIKUN ZHANG, YUHUI ZHANG, LUCIA ZHENG, KAITLYN ZHOU & PERCY LIANG, ON THE OPPORTUNITIES AND RISKS OF FOUNDATION MODELS, CTR. RSCH. ON FOUND. MODELS, INST. HUMAN-CENTERED AI, STAN. U. 3 (July 12, 2022), https://arxiv.org/pdf/2108.07258.pdf [https://perma.cc/X3JH-JUTF]; *see also* Adam Zewe, *Solving a Machine-Learning Mystery*, MASS. INST. TECH. NEWS OFF. (Feb. 7, 2023), https://news.mit.edu/2023/large-language-models-in-context-learning-0207 [https://perma.cc/S74T-EXNT] (describing LLM's "curious phenomenon known

The rapid adoption of these foundation models has elicited concern from a variety of social, business, and governmental sources.[15] Similarly, overarching societal questions require an accurate assessment of the opportunities and risks these novel artificial intelligence technologies and their implementation create.[16]

However, legal academic treatment of AI systemic risk has generally been compartmentalized into categories like specific use cases, legal questions, or remedies.[17] This trend is understandable given

as in-context learning, in which a large language model learns to accomplish a task after seeing only a few examples—despite the fact that it wasn't trained for that task"); Thomas K. Cheng & Julian Nowag, *Algorithmic Predation and Exclusion*, 25 U. PA. J. BUS. L. 41, 44 (2023) (describing these models and acknowledging their potential for "precise customer segmentation"); Jon Turow, Palak Goel & Tim Porter, *Our View on the Foundation Model Stack*, MADRONA (Jan. 27, 2023) https://www.madrona.com/foundation-models/ [https://perma.cc/2BSV-V9UY] (providing a company taxonomy of "The Foundation Model Stack"); Mike Murphy, *What are Foundation Models*?, IBM (May 9, 2022), https://research.ibm.com/blog/what-are-foundation-models [https://perma.cc/2CPB-JBCL]. For a succinct description of traditional models that illustrates the contrast with foundation models, *see* Major Aaron Kirk, *Artificial Intelligence and the Fifth Domain*, 80 A.F. L. REV., 183, 193–94 (2019).

[15] *See, e.g.*, John D. McKinnon & Ryan Tracy, *ChatGPT Comes Under Investigation by Federal Trade Commission*, WALL ST. J. (July 13, 2023), https://www.wsj.com/articles/chatgpt-under-investigation-by-ftc-21e4b3ef [https://perma.cc/EY55-C4XV]; Jonathan Haidt & Eric Schmidt, *AI Is About to Make Social Media (Much) More Toxic*, ATLANTIC (May 5, 2023), https://www.theatlantic.com/technology/archive/2023/05/generative-ai-social-media-integration-dangers-disinformation-addiction/673940/ [https://perma.cc/H9P9-CLP7]; Ethan Dodd, *The Top US Consumer Watchdog is Worried You're Going to Fall for AI Scams or Overblown Marketing Hype*, BUS. INSIDER (Mar. 3, 2023), https://www.businessinsider.com/chatbot-ftc-chatgpt-hype-scam-fraud-ai-artificial-intelligence-2023-3 [https://perma.cc/85T5-896B]; Andrew Ross Sorkin, Ravi Mattu, Bernhard Warner, Sarah Kessler, Michael J. de la Merced, Lauren Hirsch & Ephrat Livni, *Why Lawmakers Aren't Rushing to Police A.I.*, N.Y. TIMES (Mar. 2, 2023), https://www.nytimes.com/2023/03/03/business/dealbook/lawmakers-ai-regulations.html [https://perma.cc/3XZR-35PP]; *cf.* Paul Tassi, *Artists Are Mad About Marvel's 'Secret Invasion' AI-Generated Opening Credits*, FORBES (June 21, 2023), https://www.forbes.com/sites/paultassi/2023/06/21/artists-are-mad-about-marvels-secret-invasion-ai-generated-opening-credits/?sh=5f1cf5d2e6a4 [https://perma.cc/V8QP-77N6].

[16] *See generally* Matthias C. Rillig, Marlene Ågerstrand, Mohan Bi, Kenneth A. Gould & Uli Sauerland, *Risks and Benefits of Large Language Models for the Environment*, 57 ENV'T. SCI. TECH., 3464 (2023) (considering environmental effects); John Nay, *Large Language Models as Lobbyists*, STAN. L. SCH. BLOGS (Jan. 6, 2023), https://law.stanford.edu/2023/01/06/large-language-models-as-lobbyists/ [https://perma.cc/X8XP-2Q9B]; LAURA WEIDINGER, JONATHAN UESATO, MARIBETH RAUH, CONOR GRIFFIN, PO-SEN HUANG, JOHN MELLOR, AMELIA GLAESE, MYRA CHENG, BORJA BALLE, ATOOSA KASIRZADEH, COURTNEY BILES, SASHA BROWN, ZAC KENTON, WILL HAWKINS, TOM STEPLETON, ABEBA BIRHANE, LISA ANNE HENDRICKS, LAURA RIMELL, WILLIAM ISAAC, JULIA HAAS, SEAN LEGASSICK, GEOFFREY IRVING & IASON GABRIEL, TAXONOMY OF RISKS POSED BY LANGUAGE MODELS, 21415 (FAccT June 2022), https://dl.acm.org/doi/10.1145/3531146.3533088 [https://perma.cc/24QY-E7YN].

[17] *See* Rebecca Crootof & BJ Ard, *Structuring Techlaw*, 34 HARV. J. L. & TECH. 347, 348–49 (2021) (describing the state of legal analysis of novel technologies and approving in n.5 of the interdisciplinary approach in Ryan Calo, *Artificial Intelligence Policy: A Primer and Roadmap*, 51 U.C. DAVIS L. REV. 399, 516 (2017)); Alicia Solow-Niederman, *Administering Artificial Intelligence*,

the path-dependent, ongoing iterative legal process of applying prior doctrines to new facts,[18] but it does not provide a holistic understanding of the nature of emergent systemic risk from advanced AI systems' proliferation.[19] This trend is also concerning because many academic proposals for addressing AI systemic risk, recognizing the futility of a compartmentalized approach, involve cooperation across subject matters by various kinds of parties.[20] This collaboration is necessary because AI's enablement of individual entities to act at scale and with unprecedented speed amplifies risk, even in decentralized environments.[21] Thus, a more comprehensive cartography of AI systemic risk is essential to understanding the true breadth of appropriate public policy and the nature of that policy's implementation to better safeguard individuals and institutions as technological innovation necessarily continues.

But the AI risk-reward tradeoff is not calculated in a vacuum; these incredibly useful and powerful foundation models unfortunately fall into the macro-competition between nations.[22] The United States,

93 S. CAL. L. REV. 633, 650 (2020) (indicating similar trend for governmental use of AI). Attention has recently pivoted "away from technological perspectives towards societal transformation"—causing "governance approaches [to] start to come under scrutiny." Roxana Radu, *Steering the Governance of Artificial Intelligence: National Strategies in Perspective*, 40 POL'Y SOC'Y 178, 179 (2021).

[18] Oona A. Hathaway, *Path Dependence in the Law: The Course and Pattern of Legal Change in a Common Law System*, 86 IOWA L. REV. 601, 617 (2001) (describing "sequencing path dependence" as "even when actors are rational and have well-specified preferences, the order in which alternatives are presented can significantly affect the outcome"); *see also* Scott Page, *Path Dependence*, 1 Q.J. POL. SCI. 87, 88 (2006) (describing necessary elements for path dependence).

[19] *See* Michael R. Siebecker, *The Incompatibility of Artificial Intelligence and Citizens United*, 83 OHIO ST. L. J. 1211, 1223 (2022) ("Along those lines, the meaning of AI should remain essentially contextual and tethered to discrete AI applications or component technologies . . . . Within each silo, the ethical and practical considerations of AI get independently assessed, without the need for some overarching ideational construct . . . . [S]uch a compartmentalized approach to addressing the propriety of AI's development and utilization becomes rather stilted when the celerity of technological innovation causes AI applications to overlap.").

[20] *See* Hilary Allen, *Resurrecting the OFR*, 47 IOWA J. CORP. L. 1, 25 (2021); Iris Chiu & Ernest W.K. Lim, *Managing Corporations' Risk in Adopting Artificial Intelligence: A Corporate Responsibility Paradigm*, 20 WASH. U. GLOB. STUD. L. REV. 347, 360–71 (2021); *see also* Vasiliki Koniakou, *From the "Rush to Ethics" to the "Race for Governance" in Artificial Intelligence*, 25 INFO. SYST. FRONTIERS 71, 72 (2023) (describing the expansion of AI use cases). A detailed explanation of the technical workings of machine learning AI models is not the focus of this Article; legal literature has already excellently canvassed their operation. *See* Grimm et al., *supra* note 6, at 14; Brian Haney, *Applied Natural Language Processing for Law Practice*, B.C. INTELL. PROP. & TECH. F. 1, 2–22 (2020) (thoroughly discussing transformer models and natural language processing).

[21] *See* Yesha Yadav, *Fintech and International Financial Regulation*, 53 VAND. J. TRANSNAT'L L. 1109, 1120, 1126–27 (2020).

[22] *See, e.g.*, Pablo Chavez, *Vassals vs. Rivals: The Geopolitical Future of AI Competition*, LAWFARE (Aug. 3, 2023, 9:56 AM), https://www.lawfaremedia.org/article/vassals-vs.-rivals-the-geopolitical-future-of-ai-competition [https://perma.cc/F5FX-A36E].

for one, has clearly oriented the fostering of AI capabilities from a competitive perspective.[23] Furthermore, this competitive context seems to have an impact on policy preferences towards AI, pressuring governments to prioritize innovation in both the public and private sectors.[24]

However, AI development efforts face not only the typical negative tradeoffs associated with innovation,[25] but also the AI potent mix of speed and scope that may compound undetected problems or universalize known potential social ills from algorithm use.[26] For example, AI can be leveraged to promote discrimination or "human manipulation at scale" based on provoking users through the system's individualized adaptation to users' emotional states.[27] Indeed, questions of tradeoffs in foundation models and how to prioritize competing policy interests abound.[28] Policymakers should explicate a roadmap for a meta-prioritization considering how other interested

[23] *See, e.g.*, 15 U.S.C. § 9411(a) (elucidating the "purposes" of "the National Artificial Intelligence Initiative").

[24] *See* Michael Horowitz, Elsa B. Kania, Gregory C. Allen & Paul Scharre, *Strategic Competition in an Era of Artificial Intelligence*, *in* ARTIFICIAL INTEL. INT'L SEC. (Ctr. for a New Am. Sec. July 25, 2018), https://www.cnas.org/publications/reports/strategic-competition-in-an-era-of-artificial-intelligence [https://perma.cc/9MXB-TP72] (discussing AI's impact on geopolitical competition and noting: "The key power players in AI up to this point are private sector companies, not governments. For governments to effectively harness AI technology for national security uses, they will need to be able to tap into the innovation occurring in private companies."); *cf.* Or Sharir, Barak Peleg & Yoav Shoham, *The Cost of Training NLP Models: A Concise Overview*, ARXIV 3 (Apr. 19, 2020), https://arxiv.org/pdf/2004.08900.pdf [https://perma.cc/59V9-787L] ("[S]ince [natural language processing] has substantial economic value, no cost is too high in pursuit of good performance."). Not all governments prioritize innovation as dearly. For example, the European Union's AI Act has been criticized for its expansive regulatory approach. *See* Michelle Toh, *'Serious Concerns': Top Companies Raise Alarm Over Europe's Proposed AI Law*, CNN BUS. (June 30, 2023), https://www.cnn.com/2023/06/30/tech/eu-companies-risks-ai-law-intl-hnk/index.html [https://perma.cc/54Y4-GEVS] ("They argue that the draft rules go too far, especially in regulating generative AI and foundation models, the technology behind popular platforms such as ChatGPT."). Importantly, this Article uses "governments" or "government" as a catch-all term for law-making entities.

[25] *See generally* Alex Coad, Paul Nightingale, Jack Stilgoe & Antonio Vezzani, *Editorial: the Dark Side of Innovation, Industry and Innovation*, 28 INDUS. AND INNOVATION 102, 107 (2021), https://www.tandfonline.com/doi/full/10.1080/13662716.2020.1818555 [https://perma.cc/DF2R-XLMQ] (listing harms that technological and economic progress has caused).

[26] *See id.*; Guihot et. al., *supra* note 5, at 419.

[27] Louis Rosenberg, *The Profound Danger of Conversational AI*, VENTUREBEAT (Feb. 4, 2023), https://venturebeat.com/ai/the-profound-danger-of-conversational-ai/ [https://perma.cc/2S24-2UPZ]; BOMMASANI ET AL., *supra* note 14, at 19; Rishi Bommasani, Fereshte Khani, Esin Durmus, Faisal Ladhak & Dan Jurafsky, *Inequity and Fairness*, *in* BOMMASANI ET AL., *supra* note 14, at 130–32.

[28] *See* Kathleen Creel, Dallas Card, Rose E. Wang, Isabelle Levent, Alex Tamkin, Armin W. Thomas, Lauren Gillespie, Rishi Bommasani & Rob Reich, *Ethics of Scale*, *in* BOMMASANI ET AL., *supra* note 14, at 155, 160.

parties can most effectively prioritize approaching and answering these questions—a process itself likely to have downstream effects on preferred public policy.

In response to this prioritization meta-question, this Article proposes that policymakers generalize and apply Chris Brummer and Yesha Yadav's Innovation Trilemma from the financial technology (fintech) context to foundation systems' operations across use cases. This generalization and application would streamline the relationships between the issues posed by potential AI regulatory perspectives. As insightfully summarized by Brummer and Yadav:

> [W]hen seeking to (i) provide clear rules, (ii) maintain market integrity, and (iii) encourage financial innovation, regulators can achieve, at best, two out of these three objectives. For example, if regulators prioritize market safety and clear rulemaking, they necessarily must do so through broad prohibitions, likely inhibiting . . . innovation. Alternatively, if regulators wish to encourage innovation and issue clear rules, they must do so in ways that ultimately result in simple, low-intensity regulatory frameworks, increasing risks to market integrity. Finally, if regulators look to promote innovation and market integrity, they will have to do so through a complex matrix of rules and exemptions, heightening the difficulties of compliance, international coordination and enforcement.[29]

To expand the first goal, "rules simplicity reflects that regulatory dictates should attain a level of developed expression such that they provide for certainty, predictability, and stability."[30] This intention particularly manifests in standards that regulated entities understand, can use as a reference for adapting their operations into compliance, and perceive the deterrent information in those standards.[31] This also encourages "fairness" among market players in that there are little to no regulator-specific informational advantages among those entities.[32] Next, market integrity here means that financial regulators are attentive to the robustness of the market itself, as opposed to merely classic violations of law like dishonest dealings.[33] Finally, the financial innovation portion of the framework involves regulators' desires to see new ideas and technologies facilitate honest trade with more efficiency, greater information dissemination, and more insightful understanding of that information.[34]

This Article applies a version of this framework to prioritize some of the key regulatory goals for foundation models. A subtle but

[29] Chris Brummer & Yesha Yadav, *Fintech and the Innovation Trilemma*, 107 GEO. L.J. 235, 242 (2019).

[30] *Id.* at 247.

[31] *Id.*

[32] *Id.*

[33] *Id.* at 244–45.

[34] *Id.* at 246.

important alteration to this analysis is helpful to generalize the framework from a fintech context; the second consideration, "maintain market integrity,"[35] is converted to "mitigate systemic risk."[36] This change helps abstract the Trilemma to the broad capabilities and adoption, potential or realized, of foundation models. Simultaneously, it also embraces the particular concerns regarding fundamental systemic risk, transaction stability, and efficiency in fintech's use of algorithmic processes. In fintech, the outsized harms and risk of harms arise largely from the AI system's rapid execution of its operations, leading to monitoring difficulties.[37] However, foundation models introduce a greater scale to classic unobservability concerns with copious amounts of data, such as language databases.[38] The unobservability here is a different degree than that typically characterizing legal AI literature, which usually involves the human ability to understand why a model produced the outcome it did.[39] Instead, unobservability here also involves a practical inability to detect impending risk or react to error with sufficient alacrity to avoid considerable damage from seemingly spontaneous and unexpected emergent traits and other unpredictable effects model scale may produce.[40]

This Article analyzes the abstracted Trilemma to inform relative regulatory priorities. Two premises apply here. First, nation-states, companies, and other entities around the world are competing for the multifaceted comparative benefits that improved AI systems provide.[41] In the United States, private entities pioneer much of this technological advancement in the context of research collaboration and product development.[42] AI's competitive context therefore fixes innovation as a

[35] *Id.* at 242.

[36] *See id.* at 281–82.

[37] *Id.* at 280.

[38] *See* ODSC—Open Data Science, *20 Open Datasets for Natural Language Processing*, MEDIUM (July 31, 2019), https://odsc.medium.com/20-open-datasets-for-natural-language-processing-538fbfaf8e38 [https://perma.cc/Y2MJ-MSTM].

[39] *See, e.g.*, Ignacio Cofone, *Algorithmic Discrimination Is an Information Problem*, 70 HASTINGS L.J. 1389, 1438–39 (2019).

[40] *See* Stefan Buijsman & Herman Veluwenkamp, *Spotting When Algorithms Are Wrong*, MINDS AND MACH. 2–5 (Jan. 21, 2022), https://link.springer.com/article/10.1007/s11023-022-09591-0 [https://perma.cc/5TEN-DTF2] (discussing "epistemic dependence").

[41] *See generally infra* Part II.

[42] *See* Jane Zhang & Jesse Levine, *Why AI Is Next Flashpoint in US-China Tech Rivalry*, BLOOMBERG (June 29, 2023), https://www.bloomberg.com/news/articles/2023-06-29/what-is-the-state-of-us-china-competition-in-ai [https://perma.cc/XD7S-3N8G] ("The US has mostly led the way in generative AI . . . . China is seen as further ahead in fields such as image recognition, at least when it comes to practical applications."). *Compare* DANIEL ZHANG, NESTOR MASLEJ, ERIK BRYNJOLFSSON, JOHN ETCHEMENDY, TERAH LYONS, JAMES MANYIKA, HELEN NGO, JUAN CARLOS

necessary policy goal because these entities cannot outperform other business competitors and international rivals—themselves refining more advanced systems—without ongoing technological improvement.[43] Accordingly, governmental policy preferences or politics will force regulators to promote, or at least avoid excessively disturbing, AI innovation ecosystems.[44]

Taking competition spurring innovation as a given, the overriding policy question under the generalized Trilemma now turns to whether governmental entities, and US federal entities in particular, should devote more emphasis and resources to mitigating the likelihood and impacts of AI systemic risk or instead emphasize rules that are universally understandable with minimal confusion.[45] Importantly, the kind of risk that an AI system might operate in an unforeseen manner is distinct from the downstream risk for human processes built on an assumption that an underlying AI system will run as expected—and not create magnified "harms" when it does not.[46] The program, particularly a machine learning algorithm, can refine its model through mistakes.[47] However, the second premise here is that human systems built with the assumption of AI systems operating as expected are likely fragile and, therefore, unforeseen changes disproportionally harm the former systems.[48] The scope and magnitude of harms arising from these unexpected occurrences may be substantial because of the rapid spread

NIEBLES, MICHAEL SELLITTO, ELLIE SAKHAEE, YOAV SHOHAM, JACK CLARK & RAY PERRAULT, THE ARTIFICIAL INTELLIGENCE INDEX REPORT 2022, AI INDEX STEERING COMM., INST. HUMAN-CENTERED AI, STAN. U. (A.I. Index Rep. 5th ed. 2022), https://aiindex.stanford.edu/wp-content/uploads/2022/03/2022-AI-Index-Report_Master.pdf [https://perma.cc/PVC9-N9WE], *with* Gregory Dawson, Kevin C. Desouza, & James S. Denford, *Understanding Artificial Intelligence Spending by the U.S. Federal Government*, BROOKINGS (Sept. 22, 2022), https://www.brookings.edu/blog/techtank/2022/09/22/understanding-artificial-intelligence-spending-by-the-u-s-federal-government/ [https://perma.cc/7LT2-VEE8].

[43] *See generally* Eric Schmidt, *AI, Great Power Competition & National Security*, DÆDALUS, J. AM. ACAD. ARTS & SCIS. (2022), https://www.amacad.org/publication/ai-great-power-competition-national-security [https://perma.cc/6LWW-Q6R2].

[44] *See* Jack Balkin, *The Path of Robotics Law*, 6 CAL. L. REV. CIR. 45, 52–53 (2015) (recognizing the tradeoff between an aggressive liability regime and technological development).

[45] *See* Brummer & Yadav, *supra* note 29, at 249 (weighing different policy options); Balkin *supra* note 44, at 52 ("Liability without fault is a traditional solution, but it may stifle innovation in a developing area.").

[46] *See* Balkin, *supra* note 44, at 50–52.

[47] *See generally* Bhanu Garg, Li Zhang, Pradyumna Sridhara, Ramtin Hosseini, Eric Xing & Pengtao Xie, *Learning from Mistakes—A Framework for Neural Architecture Search*, ARXIV (Jan. 14, 2022), https://arxiv.org/pdf/2111.06353.pdf [https://perma.cc/4H5L-LB3M].

[48] NASSIM NICHOLAS TALEB, ANTIFRAGILE: THINGS THAT GAIN FROM DISORDER, 12 (Random House Trade Paperback ed. 2014) (defining "fragile").

and adoption of these foundation models across a variety of use cases.[49] While a lack of legal clarity or overall simplicity may lessen the efficacy of innovation promotion or the amelioration of systemic risks, this need is simply not as urgent or valuable from a downside perspective as directly attempting to prevent near-universal system disruption.

That only two priories are functionally available under the Trilemma indicates the legal regime promoting both innovation and systemic-risk mitigation must necessarily be a rather complicated one.[50] This Article thus discusses options for ameliorating the impact of regulatory complexity in this AI context.

This Article's second Part addresses the competitive landscape and rapid development of foundation models, including large language models. Market dynamics and innovation needs inform the narrow line policy should walk to foster AI innovation policy for foundation models. The third Part considers the nature of systemic risks generally; canvasses high-level perspectives and examples for how foundation models' permeation can contribute to greater systemic risk; and considers the definitional inevitability of something going wrong with inductive, probability-based systems. The fourth Part considers a sample of various AI regulatory proposals that shift risk and critiques their utility within the competition context, finishing by echoing perspectives that mitigation of harm in the face of assured eventual failure should be the legal focus for foundation models. This approach also permits a simpler articulation of legal principles. The fifth Part summarizes and concludes.

II. COMPETITION FOR AI INNOVATION'S ADVANTAGES IS ALREADY A GIVEN

This Part establishes that AI capabilities are already fundamentally paramount for both nation-state and companies' competitive capabilities. Much legal literature has made the worthwhile endeavor to explore continuity of prior doctrines, propounding normative proposals on which regulatory approaches may best satisfy those longstanding goals.[51] However, these analyses often do not place enough emphasis on the competitive circumstances in which AI innovation occurs. This Article makes no normative

[49] *Id.* at 6 ("Black Swans . . . are large-scale unpredictable and irregular events of massive consequence . . . ."); *see also* Solow-Niederman, *supra* note 17, at 661 (describing applicable breadth of subject matter).

[50] *See* Brummer & Yadav, *supra* note 29, at 249.

[51] *See, e.g.*, Karni A. Chagal-Feferkorn, *How Can I Tell if My Algorithm Was Reasonable?*, 27 MICH. TECH. L. REV. 213, 218–22 (2021).

determinations regarding the nature of this competition—it is sufficient for the present argument to accept its existence. Certainly, competition on the international stage and in markets is nothing new, but foundation models' reach, speed, and potential impact in an increasingly interconnected world has made technology and defense experts concerned that AI systems "will be weapons of first resort in future conflicts"[52] that can fundamentally damage a country's ability to distribute goods and services.[53] Moreover, earlier forms of machine learning already have ushered in a material impact across businesses, but foundation models outpace prior AI systems in terms of their potential across imaginable use cases.[54] Thus, this Article makes the non-normative observation that, unlike with typical technological innovations in the past, policymakers should particularly consider the competitive environment for use of foundation AI models to understand where innovation policy and regulation of those models will lead.

A. The Geopolitical Competitive Context

Countries, including the United States, have plainly confronted AI capabilities as a resource to acquire for competition with other nations or groups.[55] Even outside of direct AI applications, such technologies may have considerable downstream economic effects that

[52] FINAL REPORT, NAT'L SEC. COMM'N ON ARTIFICIAL INTEL. 1 (Mar. 2021), https://www.nscai.gov/wp-content/uploads/2021/03/Full-Report-Digital-1.pdf [https://perma.cc/R3E3-C8ML].

[53] 2022 NATIONAL DEFENSE STRATEGY OF THE UNITED STATES OF AMERICA, U.S. DEP'T DEF. 6 (Mar. 2023), https://media.defense.gov/2022/Oct/27/2003103845/-1/-1/1/2022-NATIONAL-DEFENSE-STRATEGY-NPR-MDR.PDF [https://perma.cc/SS9E-2SC6].

[54] Jackie Wiles, *Beyond ChatGPT: The Future of Generative AI for Enterprises*, GARTNER (Jan. 26, 2023), https://www.gartner.com/en/articles/beyond-chatgpt-the-future-of-generative-ai-for-enterprises [https://perma.cc/HAD2-TZ9T].

[55] *See, e.g.*, *International Competition Over Artificial Intelligence*, 28 STRATEGIC COMMENTS, COMMENT 11, 2–3, (June 17, 2022) ("Geopolitical analyses of AI often assume a 'race' between countries, with the key question being which country will be at the forefront of developing and deploying cutting-edge applications of AI."); *see also Summary of AI Provisions from the National Defense Authorization Act 2022*, STAN. U. HUMAN-CENTERED A.I. https://hai.stanford.edu/summary-ai-provisions-national-defense-authorization-act-2022 [https://perma.cc/AN24-V43G] (last visited Sept. 29, 2023) (identifying congressional defense spending allocations adjacent or relevant to AI); Kirk, *supra* note 14, at 185–86. Even outside of the directly competitive context, AI innovation will likely be a priority for governments in part for the approval of their constituents because investments in emerging technologies generally produce substantial social value. *See* Daniel E. Ho, Jennifer King, Russell C. Wald & Christopher Wan, *The Centrality of Data and Compute for AI Innovation: A Blueprint for the National Research Cloud*, 3 NOTRE DAME J. EMERGING TECH. 71, 76 (2022).

can bolster national resources.[56] As one example, the United States' congressionally mandated National Security Commission on Artificial Intelligence's final report begins with the following statement:

> "Americans have not yet grappled with just how profoundly the artificial intelligence . . . revolution will impact our economy, national security, and welfare . . . .[B]ig decisions need to be made now to accelerate AI innovation to benefit the United States and to defend against the malign uses of AI."[57]

Driving the point further, the report indicates that potential opponents are leveraging AI in opposition to the United States, which "will not be able to defend against AI-enabled threats without ubiquitous AI capabilities and new warfighting paradigms."[58]

As another example, the US Congress created:

> [T]he National AI Initiative to further coordinate and enhance Federal actions toward four objectives: (1) ensure continued U.S. leadership in AI research and development; (2) lead the world in the development and use of trustworthy AI systems in the public and private sectors; (3) prepare the present and future U.S. workforce for the integration of AI systems across all sectors of the economy and society; and (4) coordinate ongoing AI research, development, and demonstration activities among the civilian agencies, the Department of Defense, and the Intelligence Community to ensure that each informs the work of the others.[59]

Similarly, AI capabilities may be directly essential to thwarting AI threats from adversaries attempting to invade a system or disrupt its operations.[60] In sum, AI systems will likely be integral to the United States' defense strategy in the future.[61]

[56] *See* Philippe Aghion, Benjamin F. Jones, & Charles I. Jones, *Artificial Intelligence and Economic Growth* 2 (Nat'l Bureau of Econ. Rsch., Working Paper No. 23928, 2017), https://www.nber.org/system/files/working_papers/w23928/w23928.pdf [https://perma.cc/WM5X-C6HH] (© 2017 by Philippe Aghion, Benjamin F. Jones, and Charles I. Jones. All rights reserved.) (providing a brief synopsis on estimated economic impacts of wide-spread AI adoption).

[57] Eric Schmidt & Robert Work, *Letter from the Chair and Vice Chair* in NAT'L SEC. COMM'N ON A.I., FINAL REPORT 1 (2021), https://www.nscai.gov/wp-content/uploads/2021/03/Full-Report-Digital-1.pdf [https://perma.cc/6NLS-W4V5].

[58] *Id.* at 1–2.

[59] NAT'L A.I. RSCH. RES. TASK FORCE, STRENGTHENING AND DEMOCRATIZING THE U.S. ARTIFICIAL INTELLIGENCE INNOVATION ECOSYSTEM: AN IMPLEMENTATION PLAN FOR A NATIONAL ARTIFICIAL INTELLIGENCE RESEARCH RESOURCE 4 (Jan. 2023), https://www.ai.gov/wp-content/uploads/2023/01/NAIRR-TF-Final-Report-2023.pdf [https://perma.cc/4A75-SNRS]; *see also* 11 U.S.C. §§ 9401–9462 (statutory framework for National Artificial Intelligence Initiative).

[60] Casey & Lemley, *supra* note 7, at 357.

[61] *See, e.g.*, *Join DARPA to Reimagine the Future of AI for National Security*, DEF. ADVANCED RSCH. PROJECTS AGENCY (Feb. 24, 2023), https://www.darpa.mil/news-events/2023-02-24 [https://perma.cc/86US-XRR6]; David Vergun, *General Says Artificial Intelligence Will Play Important Role in Network Defense*, U.S. DEP'T. DEF. (Oct. 8, 2021), https://www.defense.gov/News/News-Stories/Article/Article/2805760/general-says-artificial-intelligence-will-play-important-role-in-network-defense/ [https://perma.cc/AGE2-24TH] (discussing "network protection" by AI).

This Article does not linger on a discussion of international geopolitical competition. Prior literature extensively considers how nations and companies invest in and compete by deploying advanced technology.[62] However, this competitive perspective suggests two key consequences. First, the scale and potential impact of international geopolitical competition, without more, render it unlikely that the United States, as a matter of law and policy, will deprioritize the advancement of its AI capabilities and those of its private-sector suppliers and potential partners. Second, this non-normative observation converts the Trilemma from a tripartite choice to a bipartite choice between (1) emphasizing mitigation of systemic risks and related harms, and (2) emphasizing the simplicity and ease of understanding and complying with a regulatory structure implementing AI policy goals.

B. The Private-Sector Competitive Context

The private sector largely drives foundation-model innovation in the United States,[63] and the federal government's strategy explicitly

[62] *See generally* Kristen E. Eichensehr & Cathy Hwang, *National Security Creep in Corporate Transactions*, 123 COLUM. L. REV. 549 (2023); Douglas W. Arner, Giuliano G. Castellano & Eriks Selga, *Financial Data Governance*, 74 HASTINGS L.J. 235, 238–39 (2023) (characterizing data as "a strategic resource . . . such as land, energy, food, water, and capital"); Tom C.W. Lin, *Business Warfare*, 63 B.C. L. REV. 1 (2022); Samar Fatima, Kevin C. Desouza, James S. Denford, & Gregory S. Dawson, *What Explains Governments Interest in Artificial Intelligence? A Signaling Theory Approach*, 71 ECON. ANALYSIS AND POL'Y 238 (2021), https://www.sciencedirect.com/science/article/pii/S0313592621000667 [https://perma.cc/996H-JX4B]; Kyle Wiggers, *AI Weekly: U.S. Agencies Are Increasing Their AI Investments*, VENTUREBEAT (Sept. 11, 2021), https://venturebeat.com/ai/ai-weekly-u-s-agencies-are-increasing-their-investments-in-ai/ [https://perma.cc/N45G-THUW]; Kristen E. Eichensehr, *The Law and Politics of Cyberattack Attribution*, 67 UCLA L. REV. 520 (2020); James S. Johnson, *Artificial Intelligence: A Threat to Strategic Stability*, 14 STRATEGIC STUD. Q. 16 (2020) (providing case study examples of AI impact on military). *But see* ZHANG ET AL., *supra* note 42, at 3 ("Despite rising geopolitical tensions, the United States and China had the greatest number of cross-country collaborations in AI publications from 2010 to 2021, increasing five times since 2010. The collaboration between the two countries produced 2.7 times more publications than between the United Kingdom and China—the second highest on the list." (emphasis omitted)); Alex Engler, *The EU and U.S. are Starting to Align on AI Regulation*, BROOKINGS, (Feb. 1, 2022), https://www.brookings.edu/blog/techtank/2022/02/01/the-eu-and-u-s-are-starting-to-align-on-ai-regulation/ [https://perma.cc/G2ND-9PYY]. *See generally* Houser & Raymond, *supra* note 13.

[63] *See* NAT'L ARTIFICIAL INTEL. RSCH. RES. TASK FORCE, *supra* note 59, at ii; *see also National Artificial Intelligence Initiative: Overseeing and Implementing the United States National AI Strategy*, NAT'L. A.I.GOV, https://www.ai.gov/strategic-pillars/innovation/#National_AI_Research_Institutes [https://perma.cc/4N34-S3JA] (last visited Sept. 29, 2023) (listing agencies hosting government-funded AI research efforts); *Artificial Intelligence at NSF*, NAT'L SC. FOUND. (Mar. 24, 2023), https://www.nsf.gov/cise/ai.jsp [https://perma.cc/9VMG-B58H] (describing 2020 and 2021 research project funding); Solow-Niederman, *supra* note 17, at 675–76 ("[T]he

notes its "fast-follower" relationship with the private technology industry.[64] Similarly, a reasonable hypothesis is that federal pressures on innovation will likely be reflected in policymakers' attitudes toward fostering cutting-edge AI innovation. Turning to the generalized Trilemma, policymakers' understanding of the corporate competitive context is useful to appreciate how policy can promote innovation in relationship to systemic risk and legal clarity.

The AI field is undoubtedly both itself experiencing innovation—particularly with the integration of exceptionally large sets of data with novel machine learning techniques[65]—and driving innovation in other industries.[66] The resulting AI systems have already upset some of the technology sector's business-model assumptions.[67] Such premises include the division between "the informer model," describing companies that "do not sell most end goods or services apart from some core subset of technology products," and the "seller-adviser model," focused on "sell[ing] an array of goods and services to consumers."[68] Rather, the increasing scope of AI technical ability will allow these technical products to instead provide informing services.

The historically declining cost of computing power,[69] plethora of basic online AI-creation educational services,[70] related explosion of data availability, and technology industry's open-source ethos have fostered

punchline is that public expenditures do not come close to the scale and scope of private-side R&D expenditures.").

[64] 2022 NAT'L DEFENSE STRATEGY, *supra* note 53, at 19; *see also* F. Warren McFarlan and Richard L. Nolan, *Why IT Does Matter*, HARV. BUS. SCH. WORKING KNOWLEDGE (Aug. 25, 2003), https://hbswk.hbs.edu/item/why-it-does-matter [https://perma.cc/A8DY-SJ4D] ("The first mover takes a risk and gains a temporary advantage . . . The fast follower is up against less risk but also has to recover lost ground."); Scott J. Shackelford, Isak Nti Asare, Rachel Dockery, Anjanette H. Raymond & Alexandra Sergueeva, *Should We Trust a Black Box to Safeguard Human Rights? A Comparative Analysis of AI Governance*, 26 UCLA J. INT'L L. & FOR. AFF. 37, 41–43 (2022) (describing U.S. government statements indicating reliance on private-sector development of AI).

[65] *See* Matt Bornstein, Guido Appenzeller & Martin Casado, *Who Owns the Generative AI Platform?*, ANDREESEN HOROWITZ (Jan. 19, 2023), https://a16z.com/2023/01/19/who-owns-the-generative-ai-platform/ [https://perma.cc/WSZ4-JUHH].

[66] Mark Fenwick, Wulf A. Kaal & Erik P.M. Vermeulen, *Regulation Tomorrow: What Happens When Technology Is Faster Than the Law?*, 6 AM. U. BUS. L. REV. 561, 565–67 (2017).

[67] *See* Magnuson, *supra* note 5, at 341–52.

[68] Van Loo, *supra* note 11, at 825–26.

[69] SHARIR ET AL., *supra* note 24, at 1.

[70] *See, e.g.*, *Basics of Machine Learning with TensorFlow*, TENSORFLOW, https://www.tensorflow.org/resources/learn-ml/basics-of-machine-learning [https://perma.cc/H6E4-QHFL] (last visited Sept. 10, 2023); Mark Grover, Miguel Maldonado, Joseph Santarcangelo & Xintong Li, *Unsupervised Machine Learning*, COURSERA, https://www.coursera.org/learn/ibm-unsupervised-machine-learning [https://perma.cc/7KXJ-AZZE] (last visited Sept. 29, 2023).

a "low barrier to entry" in the AI space generally.[71] But that is likely changing with respect to foundation models because a key differentiating asset for companies is the available data to train these models.[72] In other words, building an advanced system is not enough; a company generally needs the resources or platform to acquire data at a scale and rate to adequately train foundation models with current information.[73] Moreover, the level of compute needed to train the largest models may be so exceptional that later potential market entrants may be disincentivized from challenging foundation-model incumbents.[74]

Thus recognizing innovation as a critical priority, policymakers may not want to unjustly block or punish "early movers" that are creating new AI technologies,[75] particularly because larger companies are investing considerable resources across research and development in proprietary systems and in the startup landscape.[76] One example of the former is Google's 2014 acquisition and ongoing support of the AI research company DeepMind.[77] An example of the latter is Microsoft

[71] Grimm et al., *supra* note 6, at 20 ("Much, if not most commercial software relies, at least in part, on open-source software, even if it is not itself open-source."); *see* Chris Ré, *AI's Linux Moment: An Open-Source AI Model Love Note*, HAZY RSCH. (Jan. 30, 2023), https://hazyresearch.stanford.edu/blog/2023-01-30-ai-linux [https://perma.cc/5S3U-NS4S].

[72] *See* W. Nicholson Price II & Arti K. Rai, *Clearing Opacity Through Machine Learning*, 106 IOWA L. REV. 775, 791–92 (2021).

[73] *See id.* at 795–96.

[74] *See id.* at 792; *see also* Julien Simon, *Large Language Models: A New Moore's Law?*, HUGGING FACE (Oct. 26, 2021), https://huggingface.co/blog/large-language-models [https://perma.cc/N69L-4TLT]. A Stanford University team created a model called Alpaca that was "surprisingly small and easy/cheap to produce" that was nevertheless based on incumbent resources such as "Meta's LLaMa 7B model" and "OpenAIs text-davinci-003" for "instruction data." Rohan Taori, Ishaan Gulrajani, Tianyi Zhang, Yann Dubois, Xuechen Li, Carlos Guestrin, Percy Liang & Tatsunori B. Hashimoto, *Alpaca: A Strong, Replicable Instruction-Following Model*, STAN. UNIV. HUM.-CENTERED A.I., https://crfm.stanford.edu/2023/03/13/alpaca.html [https://perma.cc/T8LC-UGKN] (last visited Mar. 30, 2023). *But see* Jason Wei & Yi Tay, *Better Language Models Without Massive Compute*, GOOGLE RSCH. (Nov. 29, 2022), https://ai.googleblog.com/2022/11/better-language-models-without-massive.html [https://perma.cc/N9ZM-2TDB].

[75] *See* Tom C.W. Lin, *Artificial Intelligence, Finance, and the Law,* 88 FORDHAM L. REV. 531, 547 (2019) (describing a firm's potential for an "economic moat to shield itself from competition").

[76] *See* ZHANG ET AL., *supra* note 42, at 154 (indicating that the US private sector invested nearly $53 billion in AI during 2021); *Vijay Govindarajan*, Baruch Lev, Anup Srivastava & Luminita Enache, *The Gap Between Large and Small Companies Is Growing. Why?*, HARV. BUS. REV. (Aug. 16, 2019), https://hbr.org/2019/08/the-gap-between-large-and-small-companies-is-growing-why [https://perma.cc/VBX8-3DB9].

[77] Sam Shead, *DeepMind A.I. Unit Lost $649 Million Last Year and Had a $1.5 Billion Debt Waived by Alphabet*, CNBC (Dec. 17, 2020, 9:21 AM), https://www.cnbc.com/2020/12/17/deepmind-lost-649-million-and-alphabet-waived-a-1point5-billion-debt-.html [https://perma.cc/UW2N-4GLF]; Amit Chowdhry, *Google To Acquire Artificial Intelligence Company DeepMind*, FORBES

and OpenAI's collaboration with a venture accelerator named Neo to provide resources promoting AI-company development.[78] Nevertheless, leaders adopting nationwide policies or corporate practices advancing AI innovation impact systemic risks and may restrict the practical or politically feasible options for fairly reducing regulatory opacity; the US technology industry is often perceived as relatively consolidated, and its members have faced accusations of anticompetitive strategies and dealings.[79]

C. Potential Market Concentration in the AI Ecosystem

Unlike the Trilemma's consideration of fintech companies that were causing "an unprecedented degree of fragmentation in financial services,"[80] incumbent technology companies may benefit from the potential for foundational models' entrenchment of their competitive positions within their markets.[81] The combination of proprietary data capture and foundation models' data-crunching with other practical abilities inspires concern for incumbents' concentration.[82] Yet this concentration is not just one of market power, but also of those companies' ability to impose a broad set of preferences across subject matters by default.[83] Additionally, antitrust law has recently faced

(Jan. 27, 2014, 5:14 AM), https://www.forbes.com/sites/amitchowdhry/2014/01/27/google-to-acquire-artificial-intelligence-company-deepmind/ [https://perma.cc/UW2N-4GLF].

[78] Dina Bass, *New AI Startup Accelerator Will Partner With OpenAI, Microsoft*, BLOOMBERG (Mar. 21, 2023), https://www.bloomberg.com/news/articles/2023-03-21/openai-microsoft-to-team-up-with-ai-startups-in-neo-accelerator [https://perma.cc/2SF7-UULT].

[79] *See, e.g.*, Anca Chirita, *Abuse of Global Platform Dominance or Competition on the Merits?*, 33 LOY. CONSUMER L. REV. 1, 1 (2021); Lina M. Khan, *Amazon's Antitrust Paradox*, 126 YALE L.J. 710, 710 (2017); *see also* Mason Marks, *Biosupremacy: Big Data, Antitrust, and Monopolistic Power Over Human Behavior*, 55 U.C. DAVIS L. REV. 513, 516 (2021) ("Since 2001, five leading technology companies have avoided antitrust enforcement to complete over 600 mergers.").

[80] Brummer & Yadav, *supra* note 29, at 277.

[81] *See* Tejas N. Narechania, *Machine Learning as Natural Monopoly*, 107 IOWA L. REV. 1543, 1574 (2022) ("At least three features of machine-learning-based applications are suggestive of natural monopoly conditions (or natural monopoly effects): the costs of application development; the costs of training and optimization; and the potential for network effects."). *See generally* Rebecca Haw Allensworth, *Antitrust's High-Tech Exceptionalism*, YALE L. J. F. 589 (Jan. 18, 2021), https://www.yalelawjournal.org/pdf/AllensworthEssay_en4r12b8.pdf [https://perma.cc/P3JG-9WRW] (adopting a skeptical view towards a narrative of the relationship between antitrust law, incumbent technology firms, and innovation).

[82] *See* Narechania, *supra* note 81, at 1587–88.

[83] *See* Calo, *supra* note 17, at 424 (presenting the view that "cutting-edge AI practitioners will face even greater incentives to enter the private sphere, and [machine learning] applications will bend systematically toward the goals of profit-driven companies and not society at large. Companies will possess not only more and better information but a monopoly on its serious analysis."); *see also* Cass R. Sunstein, *Nudging: A Very Short Guide*, 37 J. CONSUMER POL'Y 583,

scrutiny for allegedly failing to restrain the influence of incumbent technology companies.[84]

An important feature of this competitive space is creating data network effects. To define, "[a] platform exhibits data network effects if, the more that the platform learns from the data it collects on users, the more valuable the platform becomes to each user."[85] Companies using advanced AI systems like LLMs can achieve substantial data network effects from more companies and customers using its platform because the company can leverage the new data that those users generate to enhance company models.[86] Indeed, scale is a logical goal for foundation-model providers because "[m]aking an improvement by [machine learning] has a high fixed cost and low marginal cost, a combination that tends to favor large firms that can spread the fixed cost over a large number of units," such as a broad population of customers.[87] Network effects and scale can similarly strengthen and

583 (2014) (describing the "nudge" concept as "liberty-preserving approaches that steer people in particular directions, but that also allow them to go their own way").

[84] *See* Lindsay Sain Jones & Tim R. Samples, *On the Systemic Importance of Digital Platforms*, 25 U. PA. J. BUS. L. 141, 203–04 (2023) ("The blind spots are substantial antitrust . . . lacks answers for social externalities, outage risks, misinformation, health impacts, and more."); Shira Ovide, *When Tech Antitrust Failed*, N.Y. TIMES (Jan. 15, 2021), https://www.nytimes.com/2021/01/15/technology/when-tech-antitrust-failed.html [https://perma.cc/W39W-VTJ9] (arguing that a government antitrust lawsuit had the indirect effect of raising product prices for consumers).

[85] Robert Gregory, Ola Henfridsson, Evgeny A. Kaganer & Harris Kryiakou, *The Role of Artificial Intelligence and Data Network Effects for Creating User Value*, 46 ACAD. MGMT. REV. 534, 535 (2021).

[86] *See* Sheen Levine & Dinkar Jain, *How Network Effects Make AI Smarter*, HARV. BUS. REV. (Mar. 14, 2023), https://hbr.org/2023/03/how-network-effects-make-ai-smarter [https://perma.cc/8DXF-HSER]; *see also* Kyle Wiggers, *Addressing Criticism, OpenAI Will No Longer Use Customer Data to Train Its Models by Default*, TECHCRUNCH (Mar. 1, 2023), https://techcrunch.com/2023/03/01/addressing-criticism-openai-will-no-longer-use-customer-data-to-train-its-models-by-default/ [https://perma.cc/7DF9-67CJ]; Uri Y. Hacohen, *Policy Implications of User-Generated Data Network Effects*, 33 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 340, 352–56 (detailing company-sourced consumer benefits from user-data network effects). *See generally* Gregory et al., *supra* note 85.

[87] S. Scott Hemphill, *Disruptive Incumbents: Platform Competition in an Age of Machine Learning*, 119 COLUM. L. REV. 1973, 1977 (2019); *see* Rich Sutton, *The Bitter Lesson*, (Mar. 13, 2019), http://www.incompleteideas.net/IncIdeas/BitterLesson.html [https://perma.cc/JN2A-PFFS] ("The biggest lesson that can be read from 70 years of AI research is that general methods that leverage computation are ultimately the most effective, and by a large margin."). *But see* Julian Michael, Ari Holtzman, Alicia Parrish, Aaron Mueller, Alex Wang, Angelica Chen, Divyam Madaan, Nikita Nangia, Richard Yuanzhe Pang, Jason Phang & Samuel R. Bowman, *What Do NLP Researchers Believe? Results of the NLP Community Metasurvey*, ARXIV 10 (Aug. 22, 2022), https://arxiv.org/pdf/2208.12852.pdf [https://perma.cc/CG7U-2VG6] (suggesting practitioner skepticism on the importance of scale).

increase by reducing transaction costs for leveraging AI platforms in other contexts.[88]

According to one view, regulators have allegedly subjected themselves to a myopic view toward these incentives by limiting those regulators' consideration of scale in their antitrust review.[89] Specifically, anticompetition enforcement efforts towards technology platforms have faced criticism for not accounting for uses of data that improve internal AI capabilities that one can deploy across an organizational structure.[90] One such critique focuses on the narrowness of the European Commission's advertising-based restrictions on Google's use of biometric information available from its FitBit acquisition.[91] For example, those restrictions did not account for if the data would be used outside of advertising practices, such as with Google's subsidiary DeepMind using the data instead to improve AI systems that could be leveraged throughout the full company, thus boosting data economies of scope and scale.[92] In short, a company with both consumer products that accumulate user data and AI systems still gains despite narrow limitations.

Breaking into the core foundation model market thus requires startups to confront three paths for data acquisition:

> [B]uild[ing] the databases themselves . . . buy[ing] the data, or . . . us[ing] 'low friction' alternatives such as content in the public domain. The last option carries perils for bias . . . . The first two are avenues largely available to big firms or institutions such as Facebook or the military.[93]

Accordingly, a proprietary source of voluminous amounts of consumer data is an incredibly difficult competitive advantage to surmount.[94] The alternative imposes limitations on distinguishing an AI product from other companies that are also using the same open-source data sets or

[88] *See* Greg Brockman, Atty Eleti, Elie Georges, Joanne Jang, Logan Kilpatrick, Rachel Lim, Luke Miller & Michelle Pokrass, *Introducing ChatGPT and Whisper APIs*, OPENAI (Mar. 1, 2023), https://openai.com/blog/introducing-chatgpt-and-whisper-apis [https://perma.cc/P7N8-RBRN].

[89] *See* Marks, *supra* note 79, at 567.

[90] *See id.*

[91] *Id.*

[92] *See* Avi Goldfarb & Daniel Trefler, *AI and International Trade, National Bureau of Economic Research* 6–8 (Nat'l Bureau of Econ. Rsch., Working Paper No. 24254, 2018) https://www.nber.org/system/files/working_papers/w24254/w24254.pdf [https://perma.cc/7HQV-YASK] (© 2018 by Avi Goldfarb and Daniel Trefler. All rights reserved.) (describing economies of scale and scope); Dan Awrey & Joshua Macey, *The Promise & Perils of Open Finance*, 40 YALE J. ON REG. 1, 38–39 (2023) (describing snowballing reinforcement of economies of scale in the information-collection space).

[93] Calo, *supra* note 17, at 424.

[94] *See id.*

poor-quality scraped information, which itself may contain copyrighted works.[95] As general collections of information that can train models eventually might hit a theoretical limit of all relevant publicly available data, the differences in an organization's proprietary data streams may become one fundamental distinction among various foundation model products.[96]

The importance of developing private and proprietary human-generated language may simply be necessary to long-term innovation and company survival in the market. Training new models from LLM-crafted website content can cause "model collapse—a degenerative process whereby, over time, models forget the true underlying data distribution, even in the absence of a shift in the distribution over time."[97] This trend causes diminished model efficacy and therefore indicates "a 'first mover advantage'" for individuals or organizations developing AI models.[98]

An aggressive regulatory policy regarding tort liability or other risk-shifting arrangements therefore might exacerbate concentration.[99] Increased expenses from regulatory requirements could prevent entrants from challenging this market, entrenching incumbent companies because algorithmic benefits from data collection cumulate over time, such that foundation model providers have a "data endowment."[100] Accordingly, users' "data portability" or

[95] *See* Nicholas Carlini, Jamie Hayes, Milad Nasr, Matthew Jagielski, Vikash Sehwag, Florian Tramèr, Borja Balle, Daphne Ippolito & Eric Wallace, *Extracting Training Data from Diffusion Models*, ARXIV 2, 6 (Jan. 30, 2023), https://arxiv.org/pdf/2301.13188.pdf [https://perma.cc/P4M5-6M7Z] ("We find that a significant number of these images fall under an explicit non-permissive copyright notice (35%). Many other images (61%) have no explicit copyright notice but may fall under a general copyright protection for the website that hosts them (e.g., images of products on a sales website).").

[96] *See* Bornstein et al., *supra* note 65; SHARIR ET AL., *supra* note 24, at 3; *see also* J. Collis & Cynthia A. Montgomery, *Competing on Resources*, HARV. BUS. REV. (July–Aug. 2008), https://hbr.org/2008/07/competing-on-resources [https://perma.cc/4Q49-E9S5] (describing the importance of "inimitable" resources).

[97] Ilia Shumailov, Zakhar Shumaylov, Yiren Zhao, Yarin Gal, Nicholas Papernot & Ross Anderson, *The Curse of Recursion: Training on Generated Data Makes Models Forget*, ARXIV 2 (May 31, 2023), https://arxiv.org/pdf/2305.17493v2.pdf [https://perma.cc/9XCH-UZHW].

[98] *Id.* at 13.

[99] *See* Seth C. Oranburg, *Encouraging Entrepreneurship and Innovation Through Regulatory Democratization*, 57 SAN DIEGO L. REV. 757, 759 (2020) (citing Mirit Eyal-Cohen, *The Cost of Inexperience*, 69 ALA. L. REV. 859, 863–64 (2018)).

[100] *See Artificial Intelligence and Machine Learning in Financial Services: Market Developments and Financial Stability Implications*, FIN. STABILITY BD. 29–30 (2017), https://www.fsb.org/wp-content/uploads/P011117.pdf [https://perma.cc/9BVP-V2PY]; Hemphill, *supra* note 87, at 1978–79; Omri Ben-Shahar, *Data Pollution*, 11 J. LEGAL ANALYSIS 104, 137 (2019) (providing the term "data endowment"); *see also* Yafit Lev-Aretz & Katherine J. Strandburg, *Privacy Regulation and Innovation Policy*, 22 YALE J. L. & TECH. 256, 294–95 (2020) ("[M]any

"interoperability" is becoming a topic of interest among scholars to alleviate anticompetitive concerns due to AI market structure.[101] Notably, some entities might have a greater appetite for legal risk than others when attempting to collect training information, forcing others to subsidize their development through aggressive and arguably illegal data aggregation.[102] Accordingly, a foundation-model provider's dominance may be path dependent by acclimating to its user base and expanding its capabilities based on incorporating user feedback.[103] This trend would further incentivize a company to expediently get to market, even with an imperfect product.[104] With these characteristics of a maturing AI industry and accompanying lopsided benefits for incumbents, a receding surge of AI startups may not be surprising. Freshly capitalized entrants into the global AI industry rose from around five hundred in 2013 to over 1,600 in 2021; those figures have since dropped to 1,392 in 2022.[105] Within the United States, the number

companies are incentivized to collect and retain as much personal data as possible simply because it might come in handy someday.").

[101] *See* Kristalina Georgiev, Federico J. Díez, Romain Duval & Daniel Schwarz, *Rising Market Power—A Threat to the Recovery?*, INT'L MONETARY FUND (Mar. 15, 2021), https://www.imf.org/en/Blogs/Articles/2021/03/15/blog-rising-market-power-a-threat-to-the-recovery [https://perma.cc/Z3WD-3VQF]; Marks, *supra* note 79, at 581; *see also* Hemphill, *supra* note 87, at 1978–79. *See generally* Gabriel Nicholas, *Taking It with You: Platform Barriers to Entry and the Limits of Data Portability*, 27 MICH. TECH. L. REV. 263 (2021) (canvassing issues relating to portability and competition). *But see* Rémy Demichelis, *Science Facing Interoperability as a Necessary Condition of Success and Evil*, ARXIV 2 (Feb. 5, 2022), https://arxiv.org/abs/2202.02540 [https://perma.cc/74KQ-UH6Q] (observing philosophically that "[w]e often would rather not consider two facts together in order to keep them apart, we would rather ignore the correlation. This separation grants not only autonomy to a sphere, but also fundamental freedom to the individuals.").

[102] *See* Gina-Gail S. Fletcher & Michelle M. Le, *The Future of AI Accountability in the Financial Markets*, 24 VAND. J. ENT. & TECH. L. 289, 303–04 (2022); *see also* Anat Lior, *The AI Accident Network: Artificial Intelligence Liability Meets Network Theory*, 95 TUL. L. REV. 1114–15 (2021) (describing a different way to view the relationship between an entity creating negative externalities and the parties that bear those externalities, the "nonreciprocal paradigm" of tort law). *See generally* Teresa Xie & Isaiah Poritz, *ChatGPT Creator OpenAI Sued for Theft of Private Data in 'AI Arms Race'*, BLOOMBERG (June 28, 2023), https://www.bloomberg.com/news/articles/2023-06-28/chatgpt-creator-sued-for-theft-of-private-data-in-ai-arms-race [https://perma.cc/DV7J-ZVVH].

[103] *See* Gregory et al., *supra* note 85, at 543; *see* Page, *supra* note 18, at 88. *See generally* Ajay Agrawal, Joshua Gans & Avi Goldfarb, *How Large Language Models Reflect Human Judgment*, HARV. BUS. REV. (June 12, 2023), https://hbr.org/2023/06/how-large-language-models-reflect-human-judgment [https://perma.cc/EX4A-JNQ2] (describing the value of user feedback for AI systems).

[104] *See* Gregory et al., *supra* note 85, at 534–35; Page, *supra* note 18, at 88.

[105] NESTOR MASLEJ, LOREDANA FATTORINI, ERIK BRYNJOLFSSON, JOHN ETCHEMENDY, KATRINA LIGETT, TERAH LYONS, JAMES MANYIKA, HELEN NGO, JUAN CARLOS NIEBLES, VANESSA PARLI, YOAV SHOHAM, RUSSELL WALD, JACK CLARK & RAYMOND PERRAULT, THE ARTIFICIAL INTELLIGENCE INDEX REPORT 2023, AI INDEX STEERING COMM., INST. HUMAN-CENTERED A.I.,

of newer financed AI companies hit a high-water mark in 2018, declined to 299 in 2021,[106] and then increased to 542 in 2022.[107] Indeed, even when a leading company releases information for other developers' use to build models, there may be (understandable) conditions for the company to leverage the data, such as prohibiting its use to create a commercial rival.[108]

This concentration has a paradoxical impact with respect to creating AI that is both useful and accommodates other social goals. While various nations have cooperated on the subject,[109] societies may also benefit if these technology incumbents also collaborated to create and maintain AI systems that avoid social harm and were broadly available on an equitable basis.[110] However, this arrangement spotlights a conflict "between cooperation and the goals of competition law, which at its core is meant to protect the very processes of rivalry between companies."[111] Thus, antitrust laws may complicate these companies' options to coordinate technology norms.

The resource-intensive data and computing scale necessary to create, train, and operate foundation AI systems might also exclude or burden different operating models from the AI industry. One example is OpenAI's transition from a nonprofit organization to a "capped-profit company": an organization that takes earnings over a certain threshold and returns them to its prior nonprofit business entity.[112] OpenAI leaders indicated the need for additional investment and talent acquisition through equity shares made the transition necessary.[113] Yet

STAN. U. 188 (A.I. Index Rep. 6th ed. 2023), https://aiindex.stanford.edu/wp-content/uploads/2023/04/HAI_AI-Index-Report_2023.pdf [https://perma.cc/3T55-FHSC].

[106] ZHANG ET AL., *supra* note 42, at 157.

[107] ARTIFICIAL INTELLIGENCE INDEX REPORT 2023, *supra* note 105, at 193.

[108] *See* Taori et al., *supra* note 74 (describing "OpenAI's text-davinci-003" free usage conditions); *Terms of use*, OPENAI, (Mar. 14, 2023), https://openai.com/policies/terms-of-use [https://perma.cc/6W4N-HS7Q] ("You may not . . . use output from the Services to develop models that compete with OpenAI . . . .").

[109] *See, e.g.*, *Our mission*, GLOB. P'SHIP ON A.I., https://gpai.ai/about/ [https://perma.cc/W8VS-9RNF] (last visited Sept. 29, 2023) (describing an international organization created in 2020 to coordinate and share research in "responsible AI" and other areas).

[110] Shin-Shin Hua & Haydn Belfield, *AI & Antitrust: Reconciling Tensions Between Competition Law and Cooperative AI Development*, 23 YALE J. L. & TECH. 415, 419 (2021); *see also* Press Release, White House, FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI (July 21, 2023), https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/ [https://perma.cc/4NVV-4V75].

[111] Hua & Beldfield, *supra* note 110, at 420.

[112] Greg Brockman & Ilya Sutskever, *OpenAI LP*, OPENAI (Mar. 11, 2019), https://openai.com/blog/openai-lp#OpenAI [https://perma.cc/2WEP-AW3L].

[113] *See id.*

commentators' criticism of this action has stemmed from that cap's upper bound being set at "[p]rofits in excess of a 100x return" on investment.[114] While this example highlights difficulties with prioritizing goals outside of investor returns, companies nevertheless emphasize building responsible models.[115] Examples of this emphasis include Amazon's Mozilla.ai or Adobe's Firefly, which Adobe says it developed exclusively and explicitly with legally permissible training data.[116]

The rush to build and implement foundation models has also opened a new competitive front among giants in a consolidated industry. Professor S. Scott Hemphill's cautionary message that antitrust enforcement should not necessarily bar technology incumbents from aggressively trying to enter each other's platform space is particularly salient for the creation and adoption of these models.[117] From this perspective, AI systems are a vehicle for large companies to attack the status quo of traditional, competitor-dominated markets. For example, news headlines have thoroughly covered Microsoft's investment in OpenAI and the incorporation of the

[114] Devin Coldewey, *OpenAI Shifts from Nonprofit to 'Capped-Profit' to Attract Capital*, TECHCRUNCH, (Mar. 11, 2019), https://techcrunch.com/2019/03/11/openai-shifts-from-nonprofit-to-capped-profit-to-attract-capital/ [https://perma.cc/C5UH-B5WU] (discussing the return cap level and responses). *But see* Paul Gompers, Will Gornall, Steven N. Kaplan & Ilya A. Strebulaev, *How Venture Capitalists Make Decisions*, HARV. BUS. REV. (Mar.–Apr. 2021), https://hbr.org/2021/03/how-venture-capitalists-make-decisions [https://perma.cc/NB4M-UK2B] (indicating how venture capitalist portfolios need massive returns on a few companies to offset other companies' performances).

[115] *See* Kyle Wiggers, *Mozilla Launches a New Startup Focused on 'Trustworthy' AI*, TECHCRUNCH (Mar. 22, 2023), https://techcrunch-com.cdn.ampproject.org/c/s/techcrunch.com/2023/03/22/mozilla-launches-a-new-startup-focused-on-trustworthy-ai/amp/ [https://perma.cc/9SK3-77S4].

[116] *See id.*; *Moz://aai: About Us*, MOZILLA.AI, https://mozilla.ai/about/ [https://perma.cc/YG2T-KANL]. *Compare* Benj Edwards, *Ethical AI Art Generation? Adobe Firefly May Be the Answer.*, ARSTECHNICA (Mar. 22, 2023), https://arstechnica.com/information-technology/2023/03/ethical-ai-art-generation-adobe-firefly-may-be-the-answer/ [https://perma.cc/5KRW-2XJJ], *with* Benj Edwards, *Artists File Class-Action Lawsuit Against AI Image Generator Companies*, ARSTECHNICA (Jan. 16, 2023), https://arstechnica.com/information-technology/2023/01/artists-file-class-action-lawsuit-against-ai-image-generator-companies/ [https://perma.cc/9WAG-ELT4] (describing a class-action lawsuit filed against other image-generation companies).

[117] *See* Hemphill, *supra* note 87, at 1974, 1997–98.

ChatGPT system across Microsoft offerings,[118] including what these changes may mean for a competitor like Google.[119]

All of this said, concentration and market dominance are hardly a given. As the foundation model and overall AI ecosystem have demonstrated furious technical innovation, no guarantee exists that these companies can keep a "durable first-mover advantage," particularly in a sector where technical innovations abound and new entrants are clawing in.[120] Additionally, incumbents may be looking for AI approaches outside of brute force scale, such as improvements to underlying approaches like transformer models.[121] Former OpenAI CEO Sam Altman conspicuously stated that "I think we're at the end of the era where it's going to be these, like, giant, giant models . . . We'll make them better in other ways."[122] Accordingly, research has pivoted toward adjusting model attributes away from scale to more productive AI systems.[123]

D. AI Innovation Norms and Needs

Having established that competition in AI is a policy priority with a complex underlying market, this Article turns to exploring aspects of the competitive environment that bolster AI innovation. A longstanding and deep culture of open collaboration between different

[118] *See, e.g.*, Tom Warren, *Microsoft Announces Copilot: The AI-Powered Future of Office Documents*, VERGE, https://www.theverge.com/2023/3/16/23642833/microsoft-365-ai-copilot-word-outlook-teams [https://perma.cc/L4XS-HCR8]; *see also Introducing the New Bing. Your AI-powered Copilot for the Web.*, MICROSOFT, https://www.microsoft.com/en-us/bing?form=MA13FJ [https://perma.cc/HR9F-HCBV] (last visited Sept. 29, 2023).

[119] *See* Nico Grant & Cade Metz, *Google Releases Bard, its Competitor in the Race to Create A.I. Chatbots*, N.Y. TIMES (Mar. 21, 2023), https://www.nytimes.com/2023/03/21/technology/google-bard-chatbot.html [https://perma.cc/HY22-ZJVA].

[120] Fernando F. Suarez & Gianvito Lanzolla, *The Half-Truth of First-Mover Advantage*, HARV. BUS. REV. (Apr. 2005), https://hbr.org/2005/04/the-half-truth-of-first-mover-advantage [https://perma.cc/B9TW-T3RF]; *see* ARTIFICIAL INTELLIGENCE INDEX REPORT 2023, *supra* note 105, at 187–90. While there might be the potential for governmental information to be a different, diffusive source of training data, currently "public entities hold far less of this valuable data when compared with the private industry." Kimberly A. Houser & John W. Bagby, *The Data Trust Solution to Data Sharing Problems*, 25 VAND. J. ENT. & TECH. L. 113, 135 (2023).

[121] *See* Will Knight, *OpenAI's CEO Says the Age of Giant AI Models Is Already Over*, WIRED (Apr. 17, 2023), https://www.wired.com/story/openai-ceo-sam-altman-the-age-of-giant-ai-models-is-already-over/ [https://perma.cc/7Z9N-VXR7].

[122] *Id.*

[123] *See, e.g.*, Marcos Treviso, Ji-Ung Lee, Tianchu Ji, Betty van Aken, Qingqing Cao, Manuel R. Ciosici, Michael Hassid, Kenneth Heafield, Sara Hooker, Colin Raffel, Pedro H. Martins, André F.T. Martins, Jessica Zosa Forde, Peter Milder, Edwin Simpson, Noam Slonim, Jesse Dodge, Emma Strubell, Niranjan Balasubramanian, Leon Derczynski, Iryna Gurevych & Roy Schwartz, *Efficient Methods for Natural Language Processing: A Survey*, ARXIV 1 (Mar. 24, 2023), https://arxiv.org/pdf/2209.00099.pdf [https://perma.cc/9ZNH-HEZV].

teams and individuals working on problems exists in AI research, but this culture is not necessarily one committed to constant assessment of product safety long-term views.[124] Rather, collaborative learning from correcting errors is essential to collective progress.[125]

Fortunately, information sharing is relatively typical when innovation takes place in the field, with company researchers often publishing key findings and studies online.[126] Similarly, software's open-source culture may be crucial for widespread innovation and ad hoc collaboration, therefore mitigating some anticompetitive risk.[127] Many large-scale, open-source resources are available to aid groups like smaller companies or teams of individuals in their creation of other more sophisticated models, such as data sets[128] or model "platform[s]" allowing developers and researchers free access.[129] For example, Meta

[124] *See* Andrew Critch & David Krueger, *AI Research Considerations for Human Existential Safety*, ARXIV 2 (June 11, 2020), https://arxiv.org/pdf/2006.04948.pdf [https://perma.cc/L2YB-5B7B] ("The field of computer science, with AI and machine learning as subfields, has not had a culture of evaluating, in written publications, the potential negative impacts of new technologies."). *See generally* Dan Hendrycks, Nicholas Carlini, John Schulman & Jacob Steinhardt, *Unsolved Problems in ML Safety*, ARXIV 13 (June 16, 2022), https://arxiv.org/pdf/2109.13916.pdf [https://perma.cc/E67G-G33W].

[125] *See* Michal Shur-Ofry, *Access-to-Error*, 34 CARDOZO ARTS & ENT. L.J. 357, 365 (2016) (citing TALEB, *supra* note 48, at 79) ("Permission is hereby granted for noncommercial reproduction of this Article in whole or in part for education or research purposes, including the making of multiple copies for classroom use, subject only to the condition that the name of the author, a complete citation, and this copyright notice and grant of permission be included in all copies.").

[126] *See, e.g.*, *Publication database*, GOOGLE RSCH., https://research.google/pubs/ [https://perma.cc/J78U-8R68] (last visited Sept. 29, 2023).

[127] *See* Hila Lifshitz-Assaf & Frank Nagle, *The Digital Economy Runs on Open Source. Here's How to Protect It.*, HARV. BUS. REV. (Sept. 2, 2021), https://hbr.org/2021/09/the-digital-economy-runs-on-open-source-heres-how-to-protect-it [https://perma.cc/Z2PE-BV3T]; Kevin Xiaoguo Zhu & Zach Zhizhong Zhou, *Lock-in Strategy in Software Competition: Open-Source Software vs. Proprietary Software*, 23 INFO. SYS. RSCH. 536, 544 (2012). *But see* Alex Engler, *How Open-Source Software Shapes AI Policy*, BROOKINGS (Aug. 10, 2021), https://www.brookings.edu/research/how-open-source-software-shapes-ai-policy/ [https://perma.cc/8FA8-CN98] ("In fact, for Google and Facebook, the open sourcing of their deep learning tools (Tensorflow and PyTorch, respectively), may have the exact opposite effect, further entrenching them in their already fortified positions.").

[128] *See, e.g.*, Alexander V. Giczy, Nicholas A. Pairolero & Andrew A. Toole, *Identifying Artificial Intelligence (AI) Invention: A Novel AI Patent Dataset*, 47 J. TECH. TRANSFER 476 (2021), https://doi.org/10.1007/s10961-021-09900-2 [https://perma.cc/XR9W-WM4S]; *Artificial Intelligence Patent Dataset*, U.S. PAT. & TRADEMARK OFF., https://www.uspto.gov/ip-policy/economic-research/research-datasets/artificial-intelligence-patent-dataset [https://perma.cc/DVH3-UVT8] (last visited Mar. 3, 2023).

[129] *See, e.g.*, *Create Production-Grade Machine Learning Models with TensorFlow*, TENSORFLOW, https://www.tensorflow.org/ [https://perma.cc/7BG2-4JWK] (last visited Sept. 29, 2023); *PyTorch*, NVIDIA, https://www.nvidia.com/en-us/glossary/data-science/pytorch/ [https://perma.cc/2HL9-3N7J] (last visited Sept. 29, 2023); *Unleashing the Potential of GANs: A Look at Popular Open-Source GAN Models*, DEFINED.AI (Feb. 20, 2023), https://www.defined.ai/blog/unleashing-the-potential-of-gans/ [https://perma.cc/WS35-JZZW]. *See generally RedPajama, a Project to Create Leading Open-Source Models, Starts by Reproducing LLaMA*

offered aspects of its LLaMa AI system to developers who wanted to use it.[130] This enabled experimental models such as Stanford's Alpaca to be created in under ten weeks for below $6 hundred.[131] This open-source culture can spill over into other industries with profound effect. For example, DeepMind and the European Molecular Biology Laboratory have combined resources to create the AlphaFold Protein Structure Database, which "provides open access to over 200 million protein structure predictions to accelerate scientific research."[132] Commentators expect such AI applications to dramatically accelerate the creation of new medical treatments.[133]

Scope of data access is not only a question of model competitive advantage but also implicates a plain need for creators of foundation models.[134] Model engineers who choose irrelevant, insufficient, or inappropriate data will produce AI that poorly accomplishes the goals its developers and users intend.[135] Nor are all goals themselves equally accomplishable; increasing accuracy requires, in part, a sufficient volume of data.[136] Fortunately, the open-source community has

Training Dataset of Over 1.2 Trillion Tokens, TOGETHER (Apr. 17, 2023), https://www.together.xyz/blog/redpajama [https://perma.cc/S95W-DRZA].

[130] Cade Metz & Mike Isaac, *In Battle Over A.I., Meta Decides to Give Away Its Crown Jewels*, N.Y. TIMES (May 18, 2023), https://www.nytimes.com/2023/05/18/technology/ai-meta-open-source.html [https://perma.cc/LEP6-ZH33].

[131] *See* Matthew Turk, *How Stanford Researchers Attempted to Make a New ChatGPT With Less Than $600*, STAN. DAILY (Apr. 2, 2023), https://stanforddaily.com/2023/04/02/how-stanford-researchers-attempted-to-make-a-new-chatgpt-with-less-than-600/ [https://perma.cc/P6NB-JPYU] (Tatsunori Hashimoto, an "Alpaca researcher," stated "I think much of the observed performance of Alpaca comes from LlaMA, and so the base language model is still a key bottleneck.").

[132] *AlphaFold Protein Structure Database*, ALPHAFOLD, https://alphafold.ebi.ac.uk/ [https://perma.cc/ME3V-QA6F] (last visited Mar. 4, 2023); *see* John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli & Demis Hassabis, *Highly Accurate Protein Structure Prediction With AlphaFold*, 596 NATURE 583 (2021), https://doi.org/10.1038/s41586-021-03819-2 [https://perma.cc/EQ9E-DC9B].

[133] *See* Vivek Subbiah, *The Next Generation of Evidence-Based Medicine*, 29 NATURE MED., 49, 50 (Jan. 16, 2023), https://www.nature.com/articles/s41591-022-02160-z [https://perma.cc/4EFX-84LN].

[134] *See* Solow-Niederman, *supra* note 17, at 688 (identifying "three broad categories of resource needs: computing power, human expertise, and data").

[135] *See* Ross P. Buckley, Dirk A. Zetzsche, Douglas W. Arner & Brian W. Tang, *Regulating Artificial Intelligence in Finance: Putting the Human in the Loop*, 43 SYDNEY L. REV. 43, 50–51 (2021).

[136] *See* Magnuson, *supra* note 5, at 355–58.

cultivated a variety of databases or corpora for developers' model training.[137] To assess progress on the effective use of this data, public contests and various public and private entities that sponsor public contests also propel the space, such as the US Department of Commerce's National Institute of Standards and Technology's Text REtrieval Conference.[138] Despite these advantages, such sharing and questions of scale present questions of systemic risk and regulatory complications.

III. SYSTEMIC RISK SHOULD BE PRIORITIZED OVER LEGAL CLARITY

Under the modified Trilemma, regulation can prioritize addressing systemic risk or legal clarity, now that policymakers' promotion of innovation is considered a given under this analysis due to international and intercompany competition.[139] The scope, impact, and potential use cases for foundation models require that legal regulation emphasize mitigation of systemic risk and its subsequent harms.[140] Legal scholars proposed the Trilemma in the analogous context of fintech, which has material similarities to the generalized, foundation-model case but also entails material dissimilarities.[141] In the finance and fintech fields, market efficiency is a key priority, but the reach of foundational models is, as previously discussed, much more pervasive.[142]

Their impact is more occluded in contrast to the high-frequency trading algorithms visibly throwing markets into chaos, as with the 2010 Flash Crash—where about $1 trillion in equity value disappeared

[137] *See, e.g.*, Peter Wayner, *22 Open Source Datasets to Boost AI Modeling*, VENTUREBEAT (Apr. 7, 2022, 11:40AM), https://venturebeat.com/data-infrastructure/22-open-source-datasets-to-fuel-your-next-project/ [https://perma.cc/35A9-CY4J].

[138] *See Text Retrieval Conference (TREC)*, NAT'L INST. STANDARDS & TECH., https://trec.nist.gov/overview.html [https://perma.cc/XZC8-H9XC] (Apr. 9, 2019, 3:31 PM); Grimm et al., *supra* note 6, at 21, 23 (describing also "[t]he Defense Advanced Research Projects Agency . . . Grand Challenge" and other contests); *see also Innovation Unleashed*, NAT'L INST. STANDARDS & TECH., https://www.nist.gov/ [https://perma.cc/3BW8-TF4D] (last visited Sept. 29, 2023). The government could stretch this concept even further by "develop[ing] public-data backed algorithms, thereby ensuring that publicly-accountable actors control a vital input for the technology." Solow-Niederman, *supra* note 17, at 689, 692–93.

[139] *See* Brummer & Yadav, *supra* note 29, at 249.

[140] *See* Gary E. Marchant & Yvonne A. Stevens, *Resilience: A New Tool in the Risk Governance Toolbox for Emerging Technologies*, 51 U.C. DAVIS L. REV. 233, 247–49 (2017) (distinguishing between reducing risk and improving resilience).

[141] *See* Brummer & Yadav, *supra* note 29, at 249.

[142] *Id.*; *see also* Brynjolfsson et al., *supra* note 3, at 19–20.

in under thirty minutes from algorithms pushing volatility.[143] Additionally, foundation models may be "multimodal," with capabilities such as image generation or robotic movement,[144] that can address use cases with considerably more reach than just market trades. This analysis therefore turns towards systemic risk to better capture the scope of potential lost positive or imposed negative externalities these models present.[145]

Legal literature on systemic risk also largely derives from financial market consideration, which is understandable given the field's history of visible and periodic eruptions such as Long-Term Capital Management's 1998 stumbles,[146] the 2008 Financial Crisis,[147] and the 2010 Flash Crash.[148] Considering this traditional lineage, Professor William Magnuson notes that the definition of systemic risk, while not settled, can be understood best through:

> [F]our factors . . . (1) the extent to which individual actors are vulnerable to rapid, adverse shocks; (2) the existence of multiple pathways for adverse shocks to spread from a single institution to others; (3) the level of asymmetric information in the market; and (4) the overall size of the market. While the presence of any one of these features in a market may not be sufficient to conclude that a market poses a systemic risk to the economy, the presence of all four should be considered a red flag.[149]

Abstracting away from a market-facing perspective, another description of such a stress-resistant system is more generalizable:

[143] Tom Lin, *The New Market Manipulation*, 66 EMORY L.J. 1253, 1260–64 (2017) (discussing the Flash Crash in detail).

[144] Gadi Singer, *Multimodality: A New Frontier in Cognitive AI*, TOWARDS DATA SCIENCE (Feb. 2, 2022), https://towardsdatascience.com/multimodality-a-new-frontier-in-cognitive-ai-8279d00e3baf [https://perma.cc/YN5P-PS54].

[145] For an excellent discussion and categorization of AI risks and their causes, *see generally* Hendrycks et al., *supra* note 124; *see also* Erik Brynjolfsson, *The Turing Trap: The Promise & Peril of Human-Like Artificial Intelligence*, 103, 107, *in* AUGMENTED EDUCATION IN THE GLOBAL AGE: ARTIFICIAL INTELLIGENCE AND THE FUTURE OF LEARNING AND WORK (Daniel Araya & Peter Marber eds. 2023) (describing beneficial and harmful externalities for human-displacing AI systems).

[146] *See generally* Paul L. Lee, *A Retrospective on the Demise of Long-Term Capital Management*, CLS BLUE SKY BLOG (Sept. 10, 2018), https://clsbluesky.law.columbia.edu/2018/09/10/a-retrospective-on-the-demise-of-long-term-capital-management/ [https://perma.cc/2QQZ-6EW2] ("[Long-Term Capital Management] was the largest hedge fund operating in the United States and its brush with death provided a preview of some of the forces that would contribute to the near collapse of the U.S. financial system in September 2008.").

[147] Daniel Schwarcz and Steven L. Schwarcz, *Regulating Systemic Risk in Insurance*, 81 U. CHI. L. REV. 1569, 1574–75 (2014).

[148] *See* Gina-Gail S. Fletcher, *Deterring Algorithmic Manipulation*, 74 VAND. L. REV. 259, 310–11 (2021). Notably, this analysis is conducted cognizant that there is some level of tradeoff between legal complexity and systemic risk. *See* J.B. Ruhl & Daniel Martin Katz, *Measuring, Monitoring, and Managing Legal Complexity*, 101 IOWA L. REV. 191, 239 (2015).

[149] William Magnuson, *Regulating Fintech*, 71 VAND. L. REV. 1167, 1189–90 (2018) (footnotes omitted).

"[R]obustness that allows the system to absorb shocks and continue serving its socially useful functions."[150] For this Article's purpose, systemic risk can similarly be understood at a high level, adopting the broad definition over the more market-focused approach.[151]

Systemic risk depends on the nature of the particular system under scrutiny. A fundamental difference exists between adaptable foundation models and the superstratum human systems (e.g., city transportation flow where many automobiles are autonomous) that depend on the assumption that those AI systems will generally work as developers and the architects of those human systems intend.[152] Systemic risk is of particular importance when competition is fierce because both private competitors and governmental authorities might deprioritize safety or downside protection.[153] Finally, AI models, even ones as powerful and dexterous as foundation models, inevitably react in unexpected and harmful ways, like Microsoft's introduction—and rapid removal—of its increasingly offensive chatbot Tay.[154] Therefore, developers' efforts to merely reduce the level of systemic risk are inadequate; efforts to mitigate harm are critical and discussed in Part IV.C. Because foundation AI models invite the creation of complex systems producing unexpected events of dramatically negative impact harm,[155] systemic risk takes precedence over simplicity of law.

[150] Allen, *supra* note 20, at 12; *see also Artificial Intelligence and Machine Learning in Financial Services: Market Developments and Financial Stability Implications*, FIN. STABILITY BD. 28 (Nov. 1, 2017), https://www.fsb.org/wp-content/uploads/P011117.pdf [https://perma.cc/6W77-359H] (The Financial Stability Board notes that "several international standards-setters have considered risks associated with algorithmic trading, as it has become a pervasive feature of markets that may, among other things, amplify systemic risk.").

[151] *Cf.* Casey & Lemley, *supra* note 7, at 296 (defining "robot" broadly for "policymakers" that do not have deep expertise with the ins and outs of the technical concept).

[152] *See* Joon Sung Park, Chris Donahue, Mina Lee, Siddharth Karamcheti, Dorsa Sadigh, Michael S. Bernstein, *Interaction, in* BOMMASANI ET AL., *supra* note 14, at 45.

[153] Hendrycks et al., *supra* note 124, at 2 ("When especially severe accidents happen, everyone loses.").

[154] Mark A. Lemley & Bryan Casey, *Remedies for Robots*, 86 U. CHI. L. REV., 1311, 1332–33 (2019); *see also* Oscar Schwartz, *In 2016, Microsoft's Racist Chatbot Revealed the Dangers of Online Conversation*, IEEE SPECTRUM (Nov. 25, 2019), https://spectrum.ieee.org/in-2016-microsofts-racist-chatbot-revealed-the-dangers-of-online-conversation#toggle-gdpr [https://perma.cc/D62Z-ZTBJ] ("[T]rolls exploited a 'repeat after me' function that had been built into Tay, whereby the bot repeated anything that was said to it on demand. . . . Tay's in-built capacity to learn meant that she internalized some of the language she was taught by the trolls, and repeated it unprompted.").

[155] *See* Michael L. Littman, Ifeoma Ajunwa, Guy Berger, Craig Boutilier, Morgan Currie, Finale Doshi-Velez, Gillian Hadfield, Michael C. Horowitz, Charles Isbell, Hiroaki Kitano, Karen Levy, Terah Lyons, Melanie Mitchell, Julie Shah, Steven Sloman, Shannon Vallor & Toby Walsh, *Gathering Strength, Gathering Storms: The One Hundred Year Study on Artificial Intelligence (AI100) 2021 Study Panel Report*, STAN. U. 77 (Sept. 16, 2021).

A. Samples of Traditional AI Models' Systemic Risks

A broad range of risk categories arise from traditional AI models. This discussion's purpose is to briefly illustrate the relevance of a broad systemic risk consideration to a policymaker's application of the abstracted Trilemma to AI generally and to contrast a subsequent discussion of the exacerbation of systemic risks with foundation AI models. This Section covers only a sliver of the relevant systemic risk typology due to its expansive breadth, but illustrations here are sufficient to demonstrate the abstracted Trilemma's utility.[156]

1. Supervision's Conflict with Speed

Speed of execution and output are defining attributes of foundation models.[157] These characteristics lead to what the Article colloquially defines a "micro"-problem of speed as involving AI systems that deviate from expectations on an operational, task-specific level and there are "macro"-problems of speed arising from the rate of innovation in AI and the "pacing problem."[158]

An algorithm's automatic and generally continuous nature exemplifies regulatory micro-problems with foundation models; unlike many organisms, algorithms do not need sleep.[159] Additionally, harm from algorithms can occur faster than humans can understand what is occurring, such as with the 2010 Flash Crash.[160] This pair of

[156] Listing systemic risks of most systems, at end, is functionally a test of one's creativity. For a thorough examination of foundation model risks, *see generally* BOMMASANI ET AL., *supra* note 14.

[157] Ahmed H. Awadallah, *AI Explainer: Foundation Models and the Next Era of AI*, MICROSOFT RSCH. BLOG (Mar. 23, 2023), https://www.microsoft.com/en-us/research/blog/ai-explainer-foundation-models-and-the-next-era-of-ai/ [https://perma.cc/86GM-BMQG].

[158] Jón Daníelsson, Robert Macrae & Andreas Uthemann, *Artificial Intelligence and Systemic Risk*, 140 J. BANKING & FIN. 1 (2022), https://doi.org/10.1016/j.jbankfin.2021.106290 [https://perma.cc/84SL-UKC3] (defining, in the financial context, a "micro problem" as "focused on day-to-day risk, such as large daily losses on individual positions, fraud and regulatory compliance" and a "macro problem" as "[l]onger term objectives, such as the solvency of key institutions, financial stability and tail risk, risks that threaten the functioning of the financial system—systemic risk"); Solow-Niederman, *supra* note 17, at 656; Guihot et al., *supra* note 5, at 421.

[159] Garrett Kenyon, *Lack of Sleep Could Be a Problem for AIs*, SCI. AM. (Dec. 5, 2020), https://www.scientificamerican.com/article/lack-of-sleep-could-be-a-problem-for-ais/ [https://perma.cc/8AFY-CTXJ] (distinguishing between "biologically realistic networks" that benefit from "an artificial analogue of sleep" versus those systems that would not need it).

[160] Fletcher, *supra* note 148, at 293–94; *see also* Andrew Critch & David Krueger, *AI Research Considerations for Human Existential Safety (ARCHES)*, ARXIV 73 (June 11, 2020), https://arxiv.org/pdf/2006.04948.pdf [https://perma.cc/RRD3-CLHH] (recognizing a tradeoff between human supervision and speed of AI task accomplishment).

characteristics seem emblematic of short-term monitoring concerns with AI.

Regulatory macro-problems with foundation models may emerge, for example, from the "pacing problem": that rapid AI development would progress past the knowledge and legal authority of a company's regulators faster than those regulators can recognize and react to new technical developments.[161] Notably, commentators offer governmental incorporation of private sector views in the course of making policy as one approach to mitigate this complication through the more efficient direction of regulatory attention.[162]

This timing issue is compounded by the Collingridge dilemma: "that at the earliest stages of development of a new technology, regulation is difficult due to a lack of information, while in the later stages the technology is so entrenched in our daily lives that there is a resistance to regulatory change from users, developers, and investors."[163] Essentially, regulators face a Goldilocks issue of trying to apply regulation that is specific enough to address a public concern without the threat of rapid irrelevancy.[164] Complicating matters further, this arrangement renders businesses' ability to anticipate regulatory requirements without fully informed collaboration and discussion with regulators a more difficult prospect.[165]

A related issue arises when the speed of actions leading to an adverse event outpaces the speed of a preexisting system that may solve the underlying problem. This trend is especially problematic for AI systems because people's inability to know and assess all potential downsides will often mean using another automated tool or system to catch errors immediately.[166] For a generalized example of this issue, financial and legal commentator Matt Levine identifies a fundamental mismatch in problem recognition and solution deployment.[167] Levine expresses incredulity at the juxtaposition of Silicon Valley Bank's

[161] Guihot et al., *supra* note 5, at 421; *see also* Ted Rall, *ChatGPT Libeled Me. Can I Sue?*, WALL ST. J. (Mar. 16, 2023, 5:57 PM), https://www.wsj.com/articles/chatgpt-libeled-me-can-i-sue-defamation-law-artificial-intelligence-cartoonist-court-lawyers-technology-14086034?mod=hp_opin_pos_2#cxrecs_s [https://perma.cc/WJ9T-MDQX]; Solow-Niederman, *supra* note 17, at 656–57.

[162] *See* Daniel J. Grimm, *Against Regulatory Disruption*, 62 JURIMETRICS J. 347, 393 (2022).

[163] Guihot et al., *supra* note 5, at 422.

[164] *Id.*

[165] *Id.* at 423.

[166] Marchant & Stevens, *supra* note 140, at 254.

[167] Matt Levine, *Silicon Valley Bank Ran Out of Money*, BLOOMBERG (Mar. 22, 2023), https://www.bloomberg.com/opinion/articles/2023-03-22/silicon-valley-bank-ran-out-of-money [https://perma.cc/SN3L-Y4V6].

technology-accessible customer interface enabling rapid withdrawal requests with the Federal Reserve's "lender-of-last-resort system . . . [which] is still stuck in a slower, more leisurely era."[168] In sum, traditional AI systems already present issues with detecting harm and proactively assigning legal requirements to avoid or mitigate harm.

2. Cybersecurity

With the modern world's widespread digitization, sprawling networks, and interconnected hardware, cybersecurity is already critical even without considering AI systems' relevance.[169] Malicious actors could purposely attack infrastructure, or virus contagion could accidently infect a system.[170] Cybersecurity intrusions have already caused considerable pecuniary harm,[171] and AI systems can seize on bad actors' maliciously planted data that damages their performance or outright produces harm.[172] Importantly, advances in AI models are not limited to benevolent uses; rather, "[machine learning] may amplify future automated cyberattacks and enable malicious actors to increase the accessibility, potency, success rate, scale, speed, and stealth of their attacks."[173]

A company is as vulnerable as its value chain or subsidiaries.[174] Therefore, a company's sole protection of its own system is no longer

[168] *See id.* (discussing Hannah Miao, Gregory Zuckerman & Ben Eisen, *How the Last-Ditch Effort to Save Silicon Valley Bank Failed*, WALL ST. J. (Mar. 22, 2023), https://www.wsj.com/articles/how-the-last-ditch-effort-to-save-silicon-valley-bank-failed-89619cb2 [https://perma.cc/Y5QE-NPCG] and noting that "[a] tech-friendly bank with a highly digitally connected set of depositors can lose 25% of its deposits in hours, which did not seem conceivable in previous eras of bank runs").

[169] *See* Lin, *supra* note 143, at 1306–07.

[170] H. Bryan Cunningham & Shauhin A. Talesh, *Uncle Sam Re: Improving Cyber Hygiene and Increasing Confidence in the Cyber Insurance Ecosystem Via Government Backstopping*, 28 CONN. INS. L.J. 1, 23–24 (2021).

[171] Lin, *supra* note 143, 1255–56 (2017) (describing a caper where hackers "manipulate[d] the markets for certain stocks" to obtain "over $100 million in illicit gains").

[172] Hendrycks et al., *supra* note 124, at 6 (describing the potential and mechanics of "backdoor . . . vulnerabilities" (internal quotation marks omitted)); Florian Tramèr, Rohith Kuditipudi & Xuechen Li, *Security and Privacy*, *in* BOMMASANI ET AL., *supra* note 14, at 105; Lin, *supra* note 75, at 540–41 (2019) (describing cyberattacks on the Securities and Exchange Commission and on a news platform's Twitter capabilities); Lin, *supra* note 143, at 1306–07 (emphasizing the importance of cybersecurity for modern capital markets).

[173] Hendrycks et al., *supra* note 124, at 11.

[174] *See* Chirantan Chatterjee & D. Daniel Sokol, *Don't Acquire a Company Until You Evaluate Its Data Security*, HARVARD BUS. REV. (Apr. 16, 2019), https://hbr.org/2019/04/dont-acquire-a-company-until-you-evaluate-its-data-security [https://perma.cc/8J9X-97JD]. *See generally* Matteo Repetto, Domenico Striccoli, Giuseppe Piro, Alessandro Carrega, Gennaro Boggia & Raffaele Bolla, *An Autonomous Cybersecurity Framework for Next-Generation Digital Service*

sufficient; corporate leaders must implement a universal cybersecurity strategy across whole connected networks, including company "counterparties and vendors."[175] Business entities need to emphasize cybersecurity in all aspects of their operations, looking to regulatory guidance for best practices when transacting with other entities.[176]

Companies can bolster their cybersecurity protections by:

> (a) investing in cybersecurity resources, including in-house expertise and training of employees; (b) having protocols in place to cooperate swiftly with other [companies], to ensure fast detection of, and responses to, these attacks, with or without involvement of regulators; and (c) building national and international systems for sharing information as well as contingency and defen[s]e planning.[177]

As mentioned in Part II.A, AI could also be developed to help assist with cybersecurity, and additional research in this particularly underappreciated use case may produce benefits relatively quickly.[178] An AI system's assistance in its own cybersecurity may be limited in novel situations, but it could improve with technical capacity for dynamic responses during attacks.[179]

Policymakers and stakeholders have attempted to encourage adoption of cybersecurity principles via a reward and punishment approach.[180] One relatively aggressive proposal is that a system owner should not be the penalized party for lapses in cybersecurity, but rather those individuals involved in actually crafting the AI program.[181] Conversely, a reward proposal would allow for favorable treatment of pre-tax investments into cybersecurity or participation in a government-sponsored insurance regime for AI model users as

Chains, 29 J. NETWORK & SYS. MGMT. 36, 36 (2021) (proposing novel security perspective for a more digitally connected market).

[175] Lin, *supra* note 75, at 544; *see also* Lior, *supra* note 102, at 1118 ("[A]s long as the black-box problem is not resolved, society will continue to view AI entities as inherently unpredictable and hence more dangerous by design.").

[176] Lin, *supra* note 143, at 1307 (recognizing "the jointly proposed improved cybersecurity standards from the Federal Reserve, the Federal Deposit Insurance Corporation, and the Office of the Comptroller of the Currency").

[177] Buckley et al., *supra* note 135, at 52–53.

[178] Hendrycks et al., *supra* note 124, at 11–12 (rating different kinds of AI issues by "Importance," "Neglectedness," and "Tractability").

[179] *See id.* at 4.

[180] Janine S. Hiller, Kathryn Kisska-Schulze & Scott Shackelford, *Cybersecurity Carrots and Sticks*, AM. BUS. L.J., (forthcoming) (manuscript at 44, 48–49), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4322819 [https://perma.cc/2UWH-VQVD].

[181] Anat Lior, *AI Entities as AI Agents: Artificial Intelligence Liability and the AI Respondeat Superior Analogy*, 46 MITCHELL HAMLINE L. REV. 1043, 1093 (2020).

described in Part IV.B.2.[182] While the need for cybersecurity is clear, the way foundational models operate might not be.

3. Opacity of Function and Data Quality

Commentary often notes that advanced machine learning techniques are essentially "black boxes" that, by nature of their particular design, do not yield much plain insight into how an AI model output was produced.[183] Two different concepts—interpretability and explainability—guide this process, though this subsection focuses on the former.[184] While debate ensues concerning appropriate definitions, for purposes here, it proves sufficient to note that the former describes how "eas[y] it is to identify cause-and-effect relationships within the system's inputs and outputs" and the latter describes "the understanding that humans achieve in terms of the internal procedures that take place while the model is training or making decisions."[185]

Foundation models may profoundly escalate interpretability issues through the extent of data used to train a model and the different analysis and weighting the model takes when creating its outputs.[186] For example, Microsoft and Nvidia collaborated to create the Megatron-Turing Natural Language Generation model, which uses "530 billion parameters,"[187] shifting the "value[s] the model can change

[182] Hiller et al., *supra* note 180, at 49–60; *see* Tom Lin, *The New Financial Industry*, 65 ALA. L. REV. 567, 616–17; Ignacio N. Cofone, *Servers and Waiters: What Matters in the Law of A.I.*, 21 STAN. TECH. L. REV. 167, 190 n.111 (2018).

[183] *See* Charlotte A. Tschider, *Beyond the "Black Box"*, 98 DENV. L. REV. 683, 700 n.102 (2021) (collecting relevant legal scholarship).

[184] Pantelis Linardatos, Vasilis Papastefanopoulos & Sotiris Kotsiantis, *Explainable AI: A Review of Machine Learning Interpretability Methods*, 23 ENTROPY 1, 2 (2021), https://www.mdpi.com/1099-4300/23/1/18 [https://perma.cc/G5RJ-RCHB] (distinguishing between the terms).

[185] *Id.* at 2–3, 5 (Providing that "[a]n interpretable model does not necessarily translate to one that humans are able to understand the internal logic of or its underlying processes. Therefore, regarding machine learning systems, interpretability does not axiomatically entail explainability, or vice versa." Additionally, presenting a useful "[t]axonomy mind-map" linking different approaches to interpretability based on a model's "purpose.").

[186] *See id.* at 6–11 (describing explainability approaches for "deep learning methods"); BOMMASANI ET AL., *supra* note 14, at 19 ("Interpretability methods at present generally are designed for interpreting and explaining the behavior of task-specific models; the nature of foundation models (i.e., the wide array of tasks these models are beneficial for and the unexpected emergent properties they acquire) introduces new challenges for interpretability research.").

[187] Ali Alvi & Paresh Kharya, *Using DeepSpeed and Megatron to Train Megatron-Turing NLG 530B, the World's Largest and Most Powerful Generative Language Model*, MICROSOFT RSCH. BLOG (Oct. 11, 2021), https://www.microsoft.com/en-us/research/blog/using-deepspeed-and-megatron-to-train-megatron-turing-nlg-530b-the-worlds-largest-and-most-powerful-generative-language-model/ [https://perma.cc/339S-2SDZv].

independently as it learns."[188] This assumes also no malicious intervention against the algorithmic assessment attempt has occurred that could complicate an observer's understanding of how a model arrived at its outputs.[189] Indeed, these models need sufficient data quality for quicker operation and greater "accuracy."[190]

Elaborating on the last point, outdated observations obscure a growing risk that a model's assumptions are deviating from reality;[191] a data-rich environment does not necessarily provide actionable information for an AI model.[192] Indeed, experts have emphasized the need for better understanding and categorization of the ways that the world may evolve to be substantially different than the information that trained the foundation model would indicate.[193]

Nor is the input data, training, and other feedback necessarily comprehensible, due to platform AI's connectivity across different programs and resources.[194] Looking again to financial markets, algorithmic actions resulting from previous, dissimilar circumstances

[188] Kyle Wiggers, *The Emerging Types of Language Models and Why They Matter*, TECHCRUNCH (Apr. 28, 2022), https://techcrunch.com/2022/04/28/the-emerging-types-of-language-models-and-why-they-matter/ [https://perma.cc/J4Y8-ESUN] ("Parameters are the parts of the model learned from historical training data and essentially define the skill of the model on a problem, such as generating text.").

[189] Fletcher, *supra* note 148, at 317 ("Specifically, researchers have shown that some of the most promising techniques for explaining and interpreting the outputs of black box algorithms can themselves be manipulated.").

[190] Gregory et al., *supra* note 85, at 541. More technically described:

> "Data quality" includes aspects of truthfulness (the degree of conformity between the recorded val[]ue and the actual value), completeness (the extent to which the recorded values exist for all observations), consistency (the degree to which the data have been measured in the same manner across cases), and timeliness (the speed by which data observations are updated in the event of change).

Id.; *see also* Eran Kahana, *A Data Stewardship Framework for Generative AI*, STAN. L. SCH. BLOGS: CODEX (Mar. 9, 2023), https://law.stanford.edu/2023/03/09/a-data-stewardship-framework-for-generative-ai/ [https://perma.cc/K26K-WYLE] (articulating "policies and procedures that are specifically designed to ensure high quality data is continuously provided" to AI models (emphasis omitted)).

[191] *See* Magnuson, *supra* note 5, at 357 ("One potential consequence of [stale data] is that bubbles could become more dramatic—machine learning algorithms could magnify momentum in particular sectors or trends, leading to eventual and catastrophic collapse. Another potential consequence . . . is that artificial financial intelligence might prevent bubbles from bursting.").

[192] Daníelsson et al., *supra* note 158, at 141.

[193] Sang Michael Xie Ananya Kumar, Rohan Taori, Tony Lee, Shiori Sagawa, Pang Wei Koh & Tatsunori Hashimoto, *Robustness to Distribution Shifts*, *in* BOMMASANI ET AL., *supra* note 14, at 111.

[194] Balkin, *supra* note 44, at 54 ("[W]e should expect that some of the most useful and widely employed robotic and AI systems will be connected to the Internet cloud. This means that these systems will not be self-contained entities, but will continually be updated by communication with other robots and AI entities, as well as centralized and decentralized sources of information.").

can harm others by roiling markets.[195] Thus, there is a risk that the inflow of data, without more, insufficiently accounts for novel or altered circumstances in the underlying process the AI models because of the weight of data the AI previously collected.[196]

Recently, scholars have focused on the relative value of data that individuals furnish in an increasingly digitized world that uses up-to-date,[197] granular information to continuously improve models as their users compete against one another.[198] These trends suggest that the "data cost" that consumers shoulder through this data transmission historically tends to grow:

> (1) when more data is collected, (2) when data is collected over longer periods of time, (3) when data is harvested deceptively, (4) when the data collected or the inferences drawn from it are sensitive, (5) when data is enriched with information from other sources, and (6) when artificial intelligence is used to refine the data."[199]

To summarize, systemic risks from AI systems can arise from people not understanding an AI's actions and an unawareness of when an AI's stimulus data has fundamentally changed.[200]

B. Foundation Models' Different and Escalated Risks

Foundation models have emerged in part from the addition of model parameters and massive, organized repositories of information that facilitate the adaptation of AI to particular needs.[201] These models "can transfer and share significant heuristics across tasks, ranging from generalizing low-level techniques that work well for one task to new scenarios all the way to directly finding meta-techniques that work

[195] Fletcher & Le, *supra* note 102, at 298.

[196] Neel Guha et al., *Vulnerabilities in Discovery Tech*, 35 HARV. J. LAW & TECH. 581, 609–10 (2022); Daníelsson et al., *supra* note 158, at 140. Separately, privacy rights are a critical issue, but one outside of the scope of this Article.

[197] Marks, *supra* note 79, at 575–78.

[198] *See* Andrei Hagiu & Julian Wright, *When Data Creates Competitive Advantage*, HARVARD BUS. REV. (Jan.–Feb. 2020), https://hbr.org/2020/01/when-data-creates-competitive-advantage [https://perma.cc/6TMJ-WV3L].

[199] Marks, *supra* note 79, at 575.

[200] *See* Balkin, *supra* note 44, at 54.

[201] Laura Weidinger, John Mellor, Maribeth Rauh, Conor Griffin, Jonathan Uesato, Po-Sen Huang, Myra Cheng Mia Glaese, Borja Balle, Atoosa Kasirzadeh, Zac Kenton, Sasha Brown, Will Hawkins, Tom Stepleton, Courtney Biles, Abeba Birhane, Julia Haas, Laura Rimell, Lisa Anne Hendricks, William Isaac, Sean Legassick, Geoffrey Irving & Iason Gabriel, *Ethical and Social Risks of Harm from Language Models*, DEEPMIND 8 (Dec. 8, 2021), https://arxiv.org/pdf/2112.04359.pdf [https://perma.cc/N6RW-AHPY].

well across numerous kinds of problems."[202] Critically, users' incorporation of these models across markets and institutional processes means that how these models operate "will thus influence actions, decisions, or policies."[203]

Many kinds of the same risks from older AI models thus persist with foundation models; however, the magnitude and scope of these risks expand with these models' practical abilities and user adoption across society.[204] As more constituents adopt these latter models for more use cases, their particular contributions to systemic risk are critical for policymakers to acknowledge because those models "will operate in environments that are broader, larger-scale, and more highly connected with more feedback loops, paving the way to more extreme events than those seen today."[205] Moreover, the general character of AI risks are naturally easier to qualitatively discuss than to numerically gauge.[206] Similarly, this Section considers some examples of these differences and their impact on systemic risks and consequent potential harm.

1. Homogenization and Monoculture

With their scale and potential for near-universal user adaptability across use cases, foundation models present material systemic risks and potential harm. First, "[f]oundation models have led to an unprecedented level of homogenization" because downstream models that are smaller or more particularized are based on adaptations of these models.[207] These circumstances allow for rapid underlying improvements to spread across new models, but they can be problematic because "all AI systems might inherit the same problematic biases of a few foundation models."[208] Again, foundation models' capabilities exhibit a vastly broader scope than those of past AI systems

[202] Yuhuai Wu, Frieda Rong, Hongyu Ren, Sang Michael Xie, Xuechen Li, Andy Shih, Drew A. Hudson & Omar Khattab, *Reasoning and Search*, *in* BOMMASANI ET AL., *supra* note 14, at 42.

[203] Neel Guha, Peter Henderson, Lucia Zheng, Mark Krass & Daniel E. Ho, *Legality*, *in* BOMMASANI ET AL., *supra* note 14, at 147.

[204] Laura Weidinger et al., *supra* note 201, at 10.

[205] Hendrycks et al., *supra* note 124, at 3; Philip Weiser, *Entrepreneurial Administration*, 97 B.U. L. REV. 2011, 2057 (2017) ("[W]e are moving into an age where networks, more than hierarchies, can better coordinate and influence behavior and adapt to changing circumstances . . . [T]he use of traditional approaches in the midst of changing circumstances can have disastrous results.").

[206] Ryan Budish, *AI's Risky Business: Embracing Ambiguity in Managing the Risks of AI*, 16 J. BUS. & TECH. L. 259, 277–78 (2021).

[207] BOMMASANI ET AL., *supra* note 14, at 5 (citations and emphasis omitted).

[208] *Id.*

and could impart the same shortcomings across typical centers of knowledge creation and across disciplines,[209] increasing the potential for contagion in crisis events.

Systemic risk from foundation models also can arise from this "algorithmic monoculture," which is "the notion that choices and preferences will become homogeneous in the face of algorithmic curation."[210] While rapid disruptions are possible when using AI, these foundational models pose a more subtle risk if algorithmic monoculture persists because they can enable less-than-ideal or harmful decisions without causing an immediate crisis that highlights its role in producing it.[211] Moreover, these risks may not be perceptible on an organization-specific level; computer scientists Jon Kleinberg and Manish Raghavan, for example, examined algorithms in multiple-company hiring and concluded each company could obtain better results from using those tools, but the algorithms overall "result in decisions that are worse on average" and therefore could have a broadly negative impact.[212]

This monoculture can also create latent-but-sudden harms through other model commonalities, such as datasets.[213] For example, "uncurated data," such as that an internet scraper bot collects from the internet wilds, can be "poisoned" by hostile actors, affecting both systems directly using that data as well as "downstream models" that are based on those systems.[214] Yet suggestions for avoiding these issues, such as rebuilding a safe version of a system's data,[215] may be resource

[209] *Id.* (acknowledging "a homogenization across research communities. For example, similar Transformer-based sequence modeling approaches are now applied to text, images, speech, tabular data, protein sequences, organic molecules, and reinforcement learning.").

[210] Jon Kleinberg & Manish Raghavan, *Algorithmic Monoculture and Social Welfare*, 118 PROCEEDINGS OF NAT'L ACAD. OF SCIS. at 1 (2021), https://www.pnas.org/doi/epdf/10.1073/pnas.2018340118 [https://perma.cc/T3TG-FVLB]; *see also* Kathleen Creel, Dallas Card, Rose E. Wang, Isabelle Levent, Alex Tamkin, Armin W. Thomas, Lauren Gillespie, Rishi Bommasani & Bob Reich, *Ethics of Scale*, *in* BOMMASANI ET AL., *supra* note 14, at 152 ("Homogenization has the potential to amplify bias; to standardize bias, compounding injustices rather than distributing them; and to amplify arbitrary exclusion." (internal citation omitted)).

[211] Kleinberg & Raghavan, *supra* note 210, at 6.

[212] *Id.* at 1.

[213] *See* Hendrycks, *supra* note 124, at 3.

[214] *Id.* at 6 ("If an adversary uploads a few carefully crafted poisoned images, code snippets, or sentences to platforms such as Flickr, GitHub or Twitter, they [sic] can inject a backdoor into future models trained on that data." (internal footnotes omitted)).

[215] *See id.* at 7.

intensive.[216] Indeed, harm can increase from data underlying models' assumptions rapidly changing in a time of crisis, meaning algorithms perform poorly in new environments and therefore accelerate crisis.[217]

The combination of algorithmic monoculture and industry concentration entails key company risk. To borrow from Professor Lawrence White, an observation of systemic risk in finance: "We will not have achieved robustness, much less antifragility, until no single financial firm is considered systemically critical or too important to close. At that point a credible promise of no bailouts can be made and kept."[218] In related vein, Professors Lindsay Sain Jones and Tim Samples support the designation and consequent heightened regulation of "systemically important technological institutions" that "pose a wide-ranging set of risks to social systems, public institutions, and human well-being" due to a potential perilously broad influence and ability to transfer costs to others.[219] This approach recognizes the importance of the "risk of failure," associated with those entities' "quasi-regulatory roles and supra-sovereign powers" that can circumvent public-sector supervision.[220] Thus, risk accompanies scale in foundation models, which the companies producing them shoulder.[221]

2. Emergent Abilities

Another critical attribute of foundation models is the development of emergent abilities, which are those "not present in smaller models but [are] present in larger models."[222] An illustration of

[216] *See* Ga Young Lee, Lubna Alzamil, Bakhtiyar Doskenov & Arash Termechy, *A Survey on Data Cleaning Methods for Improved Machine Learning Model Performance*, ARXIV 1–2, 4–5 (Apr. 13, 2021), https://arxiv.org/pdf/2109.07127.pdf [https://perma.cc/YT49-JSHV].

[217] *See* Boris Babic, I. Glenn Cohen, Theodoros Evgeniou & Sara Gerke, *When Machine Learning Goes Off the Rails*, HARVARD BUS. REV. (Jan.–Feb. 2021), https://hbr.org/2021/01/when-machine-learning-goes-off-the-rails [https://perma.cc/XG4R-YXNB] (describing fundamental changes in reality versus model assumptions via "concept drift" and "covariate shift" (emphasis omitted)).

[218] Lawrence White, *Antifragile Banking and Monetary Systems*, 33 CATO J. 471, 476 (2013), https://ciaotest.cc.columbia.edu/journals/cato/v33i3/f_0029097_23616.pdf [https://perma.cc/62JV-J3JJ]; *see also* Magnuson, *supra* note 149, at 1189–90, 1190 n.110.

[219] Jones & Samples, *supra* note 84, at 146–48, 150.

[220] *Id.* at 148 n.36, 147 (noting both that a "too big to fail" financial institution could alternatively get that designation if "the nature, scope, size, scale, concentration, interconnectedness, or mix of the activities of the firm could pose a threat to financial stability" and that the Financial Stability Oversight Council "has yet to designate a SIFI using [this] standard").

[221] *See id.* at 143.

[222] Jason Wei, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama, Maarten Bosma, Denny Zhou, Donald Metzler, Ed H. Chi, Tatsunori Hashimoto, Oriol Vinyals, Percy Liang, Jeff Dean & William Fedus, *Emergent Abilities of Large Language

this phenomenon is the sudden rapid increase in various task-related accuracy rates when operating a foundation model at scale.[223] In the future, more models will likely exhibit these traits because model breadth alone is not the determining factor for emergence. Rather, better data might reveal this trait without requiring such intensive computational assets.[224]

However, these positive emergent abilities are but one side of the coin; foundation models' unexpected operation also bring downsides.[225] These include AI models making the same logical mistakes humans do when a model expands.[226] Thus, emergence complicates the predictability of a foundation model and, therefore, introduces a distinct and material systemic risk.[227]

The scale of foundation models can also undermine individual privacy or nonpublic corporate information.[228] Even if model engineers or AI companies more generally take proactive steps to obscure data's connection to particular individuals, broader swaths of relevant data and more advanced AI techniques have in practice circumvented these

Models, ARXIV 2, 4 (Aug. 2022), https://arxiv.org/pdf/2206.07682.pdf [https://perma.cc/PB3V-YE4X] (providing exhibits illustrating that "[t]he ability to perform a task . . . is emergent when a language model achieves random performance until a certain scale, after which performance significantly increases to well-above random"); *see also* J.B. Ruhl Daniel Martin Katz, *Measuring, Monitoring, and Managing Legal Complexity*, 101 IOWA L. REV. 191, 204–06 (2015) (describing "emergence" in the legal system context).

[223] Wei et al., *supra* note 222, at 19–23 (providing graphical representation of emergence qualities and comparing models' emergent abilities across specific kinds of assignments).

[224] *Id.* at 2–3.

[225] *Id.* at 8.

[226] Ian McKenzie, Alexander Lyzhov, Michael Pieler, Alicia Parrish, Ameya Prabhu, Aaron Mueller, Najoung Kim, Sam Bowman & Ethan Perez, *Inverse Scaling Prize: Second Round Winners*, https://irmckenzie.co.uk/round2 [https://perma.cc/GNL6-ABGC] (last visited Mar. 27, 2023) (detailing, *inter alia*, Sicong Huang and Daniel Wurgaft's finding of language models' diminished performance at scale for evaluating simple logical arguments); *see also* Ian McKenzie, Alexander Lyzhov, Erjan Kalybek & Ethan Perez, *Inverse Scaling / Prize*, GITHUB, https://github.com/inverse-scaling/prize#readme [https://perma.cc/3ZB7-TUUL] (last visited Mar. 27, 2023) (putting on a "contest" to find negative model emergent traits).

[227] Wei et al., *supra* note 222, at 6–7 (Interestingly, while it is difficult to predict the parameters for which emergent properties arise, "model scale is not the singular factor for unlocking an emergent ability."); *see also* Karni A. Chagal-Feferkorn, *Am I An Algorithm or A Product? When Products Liability Should Apply to Algorithmic Decision-Makers*, 30 STAN. L. & POL'Y REV 61, 103 (2019) (noting that when how a model will work is unclear, "products liability will not necessarily contribute much to safety but will likely result in higher production costs . . . Moreover, lack of foreseeability is likely to render liability costs less predictable, and in turn again delay development or result in high costs."); Calo, *supra* note 17, at 418 (mentioning that "emergent properties . . . may pose challenges for civil liability").

[228] *See* Lev-Aretz & Strandburg, *supra* note 100, at 293 ("When information is aggregated (and, often, cross-referenced and 'enhanced' with information obtained from other sources, such as data brokers) companies can infer personal details that were not directly disclosed, often with a high level of accuracy.").

protections to reveal confidential information.[229] This risk compounds the disclosures humans make when inserting information as a prompt or any other mode of sharing confidential information with an AI system.[230] AI thus generally presents a variety of systemic risks that foundation models in particular only further compound. The speed, scale, and novelty of these broad risks call for policymakers' attention, and therefore results in the modified Trilemma's result of deprioritized legal clarity.

IV. CONFRONTING LEGAL CLARITY'S DEPRIORITIZATION

Though the "legal clarity" prong of the Trilemma is important in abstract, policymakers should deprioritize this prong as a relatively less important policy goal to the extent it does not materially impair the other priorities of innovation and systemic risk.[231] Again, competition for advanced AI systems' benefits appears to be a given, and foundation models introduce a profound depth and scope of systemic risks in a novel form. Although this last prong represents an important part of government policy towards these models and likely impacts the execution of efforts to promote innovation and diminish systemic risk, this last prong cannot be as developed as the other two. Policy specificity and granularity will be key to meeting those two priorities, thus lessening the ability to meet this third goal.

The spectrum of potential regulatory activity for foundational models generally begins on one end with a "wait-and-see" approach. This approach allows people and institutions to better understand the

[229] Steven M. Bellovin, Preetam K. Dutta & Nathan Reitinger, *Privacy and Synthetic Datasets*, 22 STAN. TECH. L. REV. 1, 15–17 (2019) (describing the oscillation over time of whether portions of genetic DNA could identify individuals).

[230] Ben Tobin, *Leaked Walmart Memo Warns Employees not to Share 'Any Information About Walmart's Business' With Chatgpt or Other AI Bots*, BUS. INSIDER (Feb. 28, 2023), https://www.businessinsider.com/walmart-warns-workers-dont-share-sensitive-information-chatgpt-generative-ai-2023-2 [https://perma.cc/KQ2R-B3NV].

[231] As stated by Professor J.B. Ruhl:

> Replacing what are perceived to be "complex" legal rules with "simple" ones to run the law-and-society system model does not necessarily produce a more adaptive law-and-society system. Dynamical systems theory shows that the surprise phenomena produced by chaos, emergence, and catastrophe can occur in systems following simple, deterministic rules of motion. Legal reform therefore misses the mark when it is aimed principally at simplifying laws; rather, the full message of complexity theory is that it is more important to aim legal reform efforts toward the factors in the law-and-society system that threaten dynamical system sustainability.

J.B. Ruhl, *Complexity Theory as a Paradigm for the Dynamical Law-and-Society System: A Wake-Up Call for Legal Reductionism and the Modern Administrative State*, 45 DUKE L.J. 849, 860 (1996).

actual risks that emerge from innovations by allotting time for those innovations and their competitive environment to settle.[232] However, regulation of an innovative, useful technology that is profoundly growing in popularity has a slim window of opportunity for governmental enactment.[233] Otherwise, such technology would become "entrenched in our daily lives that there is a resistance to regulatory change from users, developers, and investors."[234] Unsurprisingly, this live-and-let-live status quo has dissipated. Technology companies have demonstrated AI's value in a multitude of use cases, causing friction with existing law and spurring additional lawmaking. For an illustrative example of the former, the US Copyright Office removed copyright protection for AI-generated images within a comic book.[235] For one example of the latter, California requires companies to inform its Department of Motor Vehicles of autonomous car crashes.[236] Therefore, a multiplicity of policies is emerging to confront the various developer and user needs that foundation models elicit.

A. Regulation by Whom?

Legal commentary and governmental attention have produced a thoughtful panorama of suggestions and guidance for policymakers on how to wrangle AI systems causing harm,[237] but these discussions and implemented law have not reached alignment on each specific issue. On one end, people look to a broad variety of analogous circumstances for

[232] Magnuson, *supra* note 5, at 379; *cf.* Llewellyn Joseph Gibbons, *No Regulation, Government Regulation, or Self-Regulation: Social Enforcement or Social Contracting for Governance in Cyberspace*, 6 CORNELL J. L. & PUB. POL'Y 475, 476–79 (1997) (describing an unregulated internet).

[233] *See* Guihot et al., *supra* note 5, at 393–96.

[234] *Id.* at 422; *see also* Fenwick et al., *supra* note 66, at 571–72 (describing the temporal aspect of a decision to regulate).

[235] Letter from Robert Kasunic, U.S. Copyright Office, to Van Lindberg, Esq., Re: Zarya of the Dawn (Registration # VAu001480196) (February 21, 2023), https://copyright.gov/docs/zarya-of-the-dawn.pdf [https://perma.cc/6GJU-49EX]; *see also* Richard Lawler, *The US Copyright Office Says You Can't Copyright Midjourney AI-Generated Images*, THE VERGE (Feb. 22, 2023), https://www.theverge.com/2023/2/22/23611278/midjourney-ai-copyright-office-kristina-kashtanova [https://perma.cc/CH4Q-BKXY].

[236] *Autonomous Vehicle Collision Reports*, STATE OF CAL. DMV, https://www.dmv.ca.gov/portal/vehicle-industry-services/autonomous-vehicles/autonomous-vehicle-collision-reports/ [https://perma.cc/6QZG-8TSY].

[237] *See, e.g.*, Memorandum, Russell T. Vought, OFF. OF MGMT. & BUDGET, Exec. Off. of President, Memorandum for the Heads of Executive Departments and Agencies 6–12 (Nov. 17, 2020), https://www.whitehouse.gov/wp-content/uploads/2020/11/M-21-06.pdf [https://perma.cc/D65N-VEMK].

guidance.[238] Yet implicit or explicit in these analyses is the question of what entities or sources of authority will shape the contrasting (or complementary) characteristics of government-imposed foundation model regulation and of private sector interests driving regulation.

Analytically, look first to governmental regulatory policy. Presently, public regulation proposals include creating additional governmental entities, adding functions or resources to current ones, and empowering all agencies to build out their expertise through additional hiring.[239] This Article does not cover legislative options under the assumption that legislation is a particularly difficult vehicle for regulating foundation models.[240] The field is advancing quickly, and legislation is systemically inflexible to rapid modification and may not always consider on-the-ground realities.[241]

Commentators note potentially analogous experiences with the Computer Fraud and Abuse Act, which "attempts to define computer hacking and the universe of computers for which it matters, but has made a hash of it. One attempt at legislative reform, and numerous court interpretations, haven't been able to fix it in over thirty years."[242] Additionally, "overbroad" and "underinclusive" statutes can be problematic to pare back, while agency experiences and ongoing analyses of policy feedback could more rapidly drive agency policy changes.[243] The combination of inflexibility and an inherent inability to predict the future is damaging for innovation; this Subsection therefore mostly considers agency regulation.[244]

[238] *See, e.g.*, Anat Lior, *supra* note 181, at 1055–84 (canvassing thoroughly a broad variety of perspectives of legally pertinent analogues for AI systems).

[239] Calo, *supra* note 17, at 429.

[240] *See, e.g.*, *Artificial Intelligence 2023 Legislation*, NAT'L CONF. STATE LEGISLATURES (July 20, 2023), https://www.ncsl.org/technology-and-communication/artificial-intelligence-2023-legislation [https://perma.cc/P529-7XPW] (providing a table populated with summaries of state legislative approaches to AI).

[241] Casey & Lemley, *supra* note 7, at 325, 329–30, 358.

[242] *Id.* at 325.

[243] *Id.* at 325–29, 358–59 (describing benefits of agency regulation over legislation). *See generally* Neal Mollen & Aaron Ver, *Agencies Can Revise, Or Abandon, Prior Regulatory Interpretations Without Notice-And-Comment Rulemaking, Says Supreme Court*, PAUL HASTINGS (Mar. 11, 2015), https://www.paulhastings.com/insights/client-alerts/agencies-can-revise-or-abandon-prior-regulatory-interpretations-without-notice-and-comment-rulemaking-says-supreme-court [https://perma.cc/2Z8H-ZU7S].

[244] *See* Casey & Lemley, *supra* note 7, at 330 ("[A] potentially more serious risk is that the law may constrain the development of the technology itself by applying a definition written with one technology in mind to a changed world in which the line that once made sense no longer does.").

1. Government Technical Expertise

Agencies need to know the subject of their regulation to have credibility when issuing regulation, particularly in the intricate context of AI.[245] But mere agency collection of information is inadequate; agencies must collect information accurately and appropriately analyze that information to confront risk. The Financial Stability Oversight Council proves an instructive example. While created to help address financial systemic risk, the Council had "suboptimal" processes to collect knowledge and perspectives necessary for understanding that risk.[246] This impairs AI's ability as a tool for improvement because people need to be able to rely on governmental statements of safety.[247] Indeed, poorly crafted and poorly executed regulation impairs confidence in the value of regulatory imprimaturs as a signal of quality.[248]

This expertise can start internally. Governmental agencies' potential use of AI systems complicates the nature and scope of hypothetical regulatory regimes aiming to help regulate these platform models.[249] This is an area of profound importance that merits further discussion, but the concept of agency use of AI is applied here merely to illustrate the development of internal expertise and familiarity with AI in government organizations generally.

Concerning AI use in federal government, the 2020 Administrative Conference of the United States (ACUS) report indicated that particularly technical or information-focused agencies have been the more enthusiastic early adopters.[250] Examples abound, including the Office of Justice Programs using AI in twelve use cases,

[245] *See* Mason Marks, *Automating FDA Regulation*, 71 DUKE L.J. 1207, 1262 (2022); *see also* Rebecca M. Bratspies, *Regulatory Trust,* 51 ARIZ. L. REV. 575, 629–30 (2009) (discussing elements of "trust" in agency regulation); Ryan Calo & Danielle Keats Citron, *The Automated Administrative State: A Crisis of Legitimacy*, 70 EMORY L.J. 797, 844 ("Agencies should look for technical ways to enhance the expertise, discretion, and capacity for individualization that justifies committing such significant public power to bureaucratic entities in the first place.").

[246] Wulf A. Kaal, *Private Investment Fund Regulation – Theory and Empirical Evidence From 1998 to 2016*, 20 U. PA. J. BUS. L. 579, 609 (2018).

[247] Guihot et al., *supra* note 5, at 407; *see* Lin, *supra* note 75, at 545 (relatedly describing the need to have "trust and faith in the stability and reliability of the financial system").

[248] *See* Daniel E. Ho, *Does Peer Review Work? An Experiment of Experimentalism*, 69 STAN. L. REV. 1, 10 (2017).

[249] David Engstrom, Daniel E. Ho, Catherine M. Sharkey & Mariano-Florentino Cuéllar, *Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies*, ADMIN. CONF. U.S. 9 (Feb. 2020), https://law.stanford.edu/wp-content/uploads/2020/02/ACUS-AI-Report.pdf [https://perma.cc/2PVC-F4YF].

[250] *See generally id.*; *see also* Marks, *supra* note 245, at 1219–21 (summarizing and discussing the 2020 ACUS report).

the US Securities and Exchange Commission (SEC) in ten use cases, NASA in nine use cases, and the US Food and Drug Administration (FDA) in eight use cases.[251] Given the US federal government's tendency to instead hire external contractors, such results may be encouraging signs of a trend for agency internal efforts to build out AI expertise.[252] However, external experts assessed those internal efforts and found few of them to be "high in sophistication."[253]

But efforts continue. Governmental agencies could possibly gain internal, hands-on competency with AI, depending on resource availability or preexisting technical culture.[254] Agencies, though, will likely find it difficult to attract AI subject matter experts because of the novelty of the industry and potential financial rewards in a burgeoning private sector space.[255] Additionally, the product of this hands-on expertise is unlikely to be an improvement on state-of-the-art foundation models. In fact, foundation models exacerbate the need for technical knowledge and adaptive perspectives across all regulatory arms of the government.[256] Because foundation models' reach is more universal than specific models whose utilities are limited to particular subject matter, this need is particularly pronounced.[257]

Industry demand for talent imposes legitimate but substantial difficulties for a governmental body attempting to affordably carry out

[251] Engstrom et al., *supra* note 249, at 16.

[252] *Id.* at 7 ("[O]f [the] profiled use cases[,] (53%) are the product of in-house efforts by agency technologists."); *see also* Timothy DiNapoli, *A Snapshot of Government-wide Contracting for FY 2021 (Interactive Dashboard)*, U.S. GOV'T ACCOUNTABILITY OFF. (Aug. 25, 2022), https://www.gao.gov/blog/snapshot-government-wide-contracting-fy-2021-interactive-dashboard [https://perma.cc/M3TH-94VB] ("In Fiscal Year 2021, the federal government spent about $637 billion on contracts, a decrease of $54 billion from FY 2020 after adjusting for inflation.").

[253] Engstrom et al., *supra* note 249, at 7.

[254] Calo & Citron, *supra* note 245, at 845 (suggesting that agency adoption of automation is in part a result of "the chronic lack of resources best laid at the feet of the legislature or executive").

[255] Daniel Zhang, Christie Lawrence, Michael Sellitto, Russell Wald, Marietje Schaake, Daniel E. Ho, Russ Altman & Andrew Grotto, *Enhancing International Cooperation in AI Research: The Case for a Multilateral AI Research Institute*, 8, STANFORD UNIV. HUM.-CENTERED A.I. (May 2022), https://hai.stanford.edu/sites/default/files/2022-05/HAI%20Policy%20White%20Paper%20-%20Enhancing%20International%20Cooperation%20in%20AI%20Research.pdf [https://perma.cc/5T77-KSFM] ("Difficulties attracting, training, and retaining skilled AI talent also significantly limit research as countries, governments, research institutions, and even private companies compete across the scarce AI labor market."); *see also* Matthew U. Scherer, *Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies*, 29 HARV. J. L. & TECH. 353, 379–80, 384–85 (2016).

[256] *See* Calo, *supra* note 17, at 428.

[257] *See id.* at 417, 427–28.

regulatory mandates.[258] Professor Hilary Allen identifies some alternative routes for governments to obtain crucial technical talent.[259] Specifically, she recommends joining efforts with university projects; leveraging the Intergovernmental Personnel Act, which permits short-term borrowing of talent from "academic institutions, governmental bodies, and other organizations deemed eligible by the agency"; legislative authorization to sidestep typical executive-branch pay caps; and quality-of-life pitches like full-time remote work.[260]

Finally, as discussed in Section III.A.1, the pacing problem can cause problems for governmental regulators because companies can circumvent a regulation's ambit via (arguably) transforming a product by incorporating new features or replacing older ones.[261] Though potentially faster than legislation, agency action takes time, and an agency might not want to give uninformed informal guidance for fear of harm to the legitimacy of the agency's lawmaking.[262] As a result, one option for novel technologies is to follow Google's steps with the National Highway Traffic Safety Administration to address driverless vehicles. Specifically, Google simply requested exemption from particular regulations,[263] adding a wrinkle that illustrates expanded company options when dealing with a particular agency.

2. Intra- or Inter-Governmental Collaboration

Foundation models, by their nature, reach across many regulatory jurisdictions through a multitude of potential use cases. Regulators have delineated subject matter expertise, rendering unified

[258] Lin, *supra* note 143, at 1295–96; Allen, *supra* note 20, at 30–31 ("[T]he OFR will have to compete with the private sector not just for the mathematical expertise that characterized earlier generations of private sector quants, but also for new types of in-demand expertise.").

[259] *See* Allen, *supra* note 20, at 31–33.

[260] *Id.*; *see also 2021 Guide to Telework and Remote Work in the Federal Government*, U.S. OFF. OF PERSONNEL MGMT. 2, (Nov. 2021), https://www.telework.gov/guidance-legislation/telework-guidance/telework-guide/guide-to-telework-in-the-federal-government.pdf [https://perma.cc/3AU9-G65S]; *see also* Cary Coglianese & David Lehr, *Regulating by Robot: Administrative Decision Making in the Machine-Learning Era*, 105 GEO. L.J. 1147, 1164 (2017) (describing the FDA's "five-year collaborative research agreement with the Massachusetts Institute of Technology . . . focusing on artificial intelligence, advanced statistical machine learning and data mining methods." (internal quotation marks omitted)).

[261] *See* Casey & Lemley, *supra* note 7, at 329 (discussing how regulatory "[u]nderinclusiveness" can occur as a practical matter).

[262] *See* Calo & Citron, *supra* note 245, at 835 ("Agencies deserve the power they possess based on their expertise, flexibility, and nimbleness. This is true not only at a pragmatic level, but also at the level of first principles.").

[263] Casey & Lemley, *supra* note 7, at 333.

policy coordination difficult.[264] Thus, if multiple agencies in the United States regulate the results of AI innovation, then they will need to collaborate on some level,[265] at least to the extent necessary to avoid political embarrassment and contradictory regulatory requirements. Critically, agencies must increase information flow and reduce communication barriers to the greatest extent reasonably possible to appropriately assess risk across society.[266]

Some agencies might not realize the extent to which their oversight mandates will overlap because of the growing ubiquity of AI.[267] Considerations accounting for internal agency workforce operations also will require substantial agency leadership to surmount. As one commentary states, "[i]n many cases, new leaders of administrative agencies are 'captured' by the bureaucracy, encouraged to accept the traditional modes of operation as a given, and" are disincentivized from engaging in experimental or dramatically atypical activities.[268] Regulatory myopia from viewing "'success' in negative terms, as in the avoidance of catastrophe,"[269] can meaningfully restrict an agency's incentives for flexibility. Cultural norms defining an agency's mission in that manner likely require substantial leadership to overcome.

Intragovernmental conflicts may also give rise to what Professors Casey and Lemley describe as "zero-sum battles over regulatory authority,"[270] manifesting in failure to share information and cooperate. The practical limitations on the Office of Financial

[264] *See, e.g.*, Fletcher, *supra* note 148, at 266 (critiquing the "disjointed and inconsistent" approaches to algorithms by the Securities and Exchange Commission and the Commodities Futures Trading Commission).

[265] *See* Jody Freeman & Jim Rossi, *Agency Coordination in Shared Regulatory Space*, 125 HARV. L. REV. 1131, 1156–78 (2012).

[266] Allen, *supra* note 20, at 33–35 (identifying memoranda of understanding as likely an appropriate mechanism for aligning agency goals, "sequencing of interagency projects," promoting job exchange programs with other regulatory entities, and coordinating work towards "a real-time data reporting system that can be used by both the regulatory agencies and the private sector").

[267] *See* Van Loo, *supra* note 11, at 875–82.

[268] Weiser, *supra* note 205, at 2077; *see also* Gillian Metzger & Kevin Stack, *Internal Administrative Law*, 115 MICH. L. REV. 1239, 1253 (2017). *See generally* Allison M. Whelan, *Executive Capture of Agency Decisionmaking*, 75 VAND. L. REV. 1787, 1805 (2022) (articulating the notion of "internal agency capture," which is when "government actors such as the President or other White House officials exert undue influence over agencies and cause them to make decisions contrary to their missions").

[269] Fenwick et al., *supra* note 66, at 575; *see* Weiser, *supra* note 205, at 2028; Jody Freeman, *The Private Role in Public Governance*, 75 N.Y.U.L. REV. 543, 671 (2000) ("Indeed, public/private engagement may enhance state power while simultaneously augmenting private power. Through contract with private actors, for example, agencies may extend their influence to matters and actors that they could not otherwise lawfully reach.").

[270] Casey & Lemley, *supra* note 7, at 337.

Research (OFR) aptly showcase this phenomenon. The OFR faced both Treasury Department preemptive pushes for the OFR to be included within that Department and other regulators' apparent avoidance of information sharing.[271]

A good solution for this coordination problem may be presidential or other authoritative selection of metrics to determine an aggregate evaluation or measurement method. Such metrics would provide easily comprehensible signals for leadership across a spectrum of agencies to understand relative benefits of collective regulatory strategies.[272] Indeed, the "Interagency Committee" authorized to facilitate AI policy under 15 U.S.C. § 9413 may be a useful vehicle for producing proposals on evaluation methods.[273] Though a full assessment of this tactic is outside the scope of the present analysis, it illustrates that formalizable approaches to coordination and information dissemination exist across governmental entities with respect to AI policy.

3. Risks of Regulatory Capture and Arbitrage

Agency capture and regulatory arbitrage are key factors that policymakers should strive to reduce when regulating foundation models. Regulatory capture occurs "when agencies consistently adopt regulatory policies favored by regulated entities."[274] First, any agency substantially regulating AI may be particularly susceptible to regulatory capture. The intellectual and financial opportunities in the private sector have been particularly exceptional and may induce individuals working in governmental agencies to enter a regulated company.[275] Next, agencies, legislatures, and other law-creation or enforcement entities may be comprised of people who, intentionally or

[271] Allen, *supra* note 20, at 7–9 (describing Treasury pressure to prevent OFR's establishment as an independent agency and OFR's conflict with the SEC "over the OFR's research on systemic risks in the asset management industry").

[272] *See* Cary Coglianese, *Measuring Regulatory Performance: Evaluating the Impact of Regulation and Regulatory Policy,* 18–23 (OECD Expert Paper No. 1, 2012), https://www.oecd.org/regreform/regulatory-policy/1_coglianese%20web.pdf [https://perma.cc/94U6-PSVC].

[273] *See* 15 U.S.C. § 9413(a).

[274] Sidney Shapiro, *Blowout: Legal Legacy of the Deepwater Horizon Catastrophe: The Complexity of Regulatory Capture: Diagnosis, Causality, and Remediation*, 17 ROGER WILLIAMS U. L. REV. 221, 224 (2012).

[275] *See e.g.*, Cade Metz, *A.I. Researchers Are Making More Than $1 Million, Even at a Nonprofit*, N.Y. TIMES (Apr. 19, 2018), https://www.nytimes.com/2018/04/19/technology/artificial-intelligence-salaries-openai.html [https://perma.cc/T94R-RWS2].

otherwise, incorporate their own preferences into policy.[276] This simple fact that people comprise agencies introduces the concern of a form of agency capture called "cultural capture," describing how a regulator staff's social preferences interfere with policy actions because of the value of being perceived as a member of, or associated with, "an elite community."[277] This aspect of public policy influence is particularly pertinent to AI regulation because technology expertise tends to gather in particular geographic communities such as the San Francisco Bay Area or the Washington-Boston corridor.[278]

Capture can manifest through passive acceptance of industry practices.[279] This is problematic in high-risk scenarios, where thorough regulatory presence is key for resolving collective action failures or comprehensive risk mitigation. Though less of a concern in a rule-based system, a standard-based system would usher in more informal influence through considerable regulator "discretion" that allows individual preferences or idiosyncrasies to have an impact on results.[280]

Cultural capture is not the sole concern when envisioning an improved regulatory AI framework. Another threat to effective regulation is that of legal endogeneity. This phenomenon occurs when private sector "compliance professionals . . . have significant power to define what the law means in practice" by "fram[ing] the law in accordance with managerial values like operational efficiency and reducing corporate risk."[281] This perspective is then later adopted by "judges and policymakers . . . as paradigms of best practices or as evidence for an affirmative defense or safe harbor."[282] In short,

[276] Ganesh Sitaraman, *The Regulation of Foreign Platforms*, 74 STAN. L. REV. 1073, 1094 (2022).

[277] *Id.* at 1094–95 (quoting James Kwak, *Cultural Capture and the Financial Crisis*, *in* PREVENTING REGULATORY CAPTURE: SPECIAL INTEREST INFLUENCE AND HOW TO LIMIT IT, 71, 78–79, 96 (Daniel Carpenter & David A. Moss eds., 2014)).

[278] *See* Mark Muro & Sifan Liu, *The Geography of AI: Which Cities Will Drive the Artificial Intelligence Revolution?*, METRO. POL'Y PROGRAM AT BROOKINGS, 13 (Sept. 2021), https://www.brookings.edu/wp-content/uploads/2021/08/AI-report_Full.pdf [https://perma.cc/2E83-GSME]; *see also* Bhaskar Chakravorti, Ajay Bhalla, Ravi Shankar Chaturvedi & Christina Filipovic, *50 Global Hubs for Top AI Talent*, HARVARD BUS. REV. (Dec. 21, 2021), https://hbr.org/2021/12/50-global-hubs-for-top-ai-talent [https://perma.cc/WR2B-2T8K]; Goldfarb & Trefler, *supra* note 92, at 9; Georg Rilinger, *Who Captures Whom? Regulatory Misperceptions and the Timing Of Cognitive Capture*, 17 REGU. & GOVERNANCE, 43, 43–44 (2023).

[279] *See* Allen, *supra* note 20, at 4 (prescribing OFR skepticism of prevailing methods of estimating rink in industry).

[280] *See* Sitaraman, *supra* note 276, at 1095.

[281] Ari Ezra Waldman, *Privacy Law's False Promise*, 97 WASH. U. L. REV. 773, 776 (2020) (discussing legal endogeneity in the privacy context).

[282] *Id.* at 777.

regulated entities define the scope of what behavior is legal; compliance acts without substance become "substantive."[283] Here, regulator collaboration with its regulated industry must not mean ceding perspectives or opinions to private industry's issue framing. Rather, regulators must also find reliable and less biased sources of expertise.[284]

While it may be financially or reputationally costly, regulated firms may pursue regulatory capture or influence as a matter of reducing uncertainty and its resulting costs.[285] This incentivization is particularly pronounced when those firms are on the outer edge of product innovation. Some evidence suggests that "regulatory co-creation," an "effort to work collaboratively with regulators, rather than try to co-opt them" might be a useful option for firms.[286] This communicative and informal approach may have an overall positive impact on successfully encouraging regulated companies to comply with requirements. Regulatory co-creation may also help companies and regulators identify feasible solutions and the requisite scope of regulation.[287]

B. Reducing Risks Through Incentive Shifting

This Section samples some proposals for producers of AI systems to accept a greater share of the risks presented by the public's use of those systems. This line of inquiry is important for all AI models, but some proposals—though eminently worthwhile—diminish the influence of the foundation model competitive context or inherent emergent abilities' unpredictability. This approach adheres to the perspective that law generally should not be beholden to competitive pressures when protecting people from harm. From a descriptive approach, though, this perspective might not adequately account for policymakers' countervailing pressure with respect to national competitiveness.[288] Therefore, a recognition of policy trade-offs between reducing imposition of risk on developers and innovation is worthwhile.

[283] *See id.* at 776.

[284] Calo, *supra* note 17, at 428 ("When the state does not have its own experts, it must either rely on the self-interested word of private firms (or their proxies) or experience a paralysis of decision and action that ill-serves innovation.").

[285] *See* Cheng Gao & Rory McDonald, *Shaping Nascent Industries: Innovation Strategy and Regulatory Uncertainty in Personal Genomics*, 36 ADMIN. SCI. Q., 915, 942 (2022).

[286] *Id.* at 944–45.

[287] *Id.* at 945, 947; Magnuson, *supra* note 5, at 378 (describing how the "precautionary principle" of basing regulation on an overwhelming preference in favor of preventing risk would impair "innovation").

[288] *See* Magnuson, *supra* note 5, at 358–59.

1. Tort Burdens

In a rapidly developing market with large incumbent firms, startups may face stark trade-offs.[289] To succeed in the competitive AI environment, they may take a high-risk, high-reward chance on untested technology, essentially passing negative externalities for their development of AI systems to end users or data sources.[290] Entrepreneurship builds its vaunted "minimum viable product" development approach, "design-centered thinking," and agile project planning methodologies on the idea that mistakes and iterations will occur.[291] A customer user base's experiences then function to provide valuable input on how their needs interface with a company's offering.[292] However, this common product development approach may fundamentally conflict with ex ante regulatory approaches; those approaches would increase the cost of each product iteration and make it more difficult for companies to find the market for new products.[293]

Much of the legal AI literature therefore appropriately considers shifting incentives to AI developers to internalize the costs and risks associated with AI systems' use.[294] The law generally incentivizes product producers to shoulder the burden of constructing a product as adequately safe as the "least cost avoider";[295] indeed, products liability

[289] *See* Tom Eisenmann, *Why Start-ups Fail,* HARVARD BUS. REV. (May–June 2021), https://hbr.org/2021/05/why-start-ups-fail [https://perma.cc/6UU3-9JJ6] (describing dire startup success rates); *see also* Chagal-Feferkorn, *supra* note 227, at 78–80 (describing "deterrence" purpose for requiring a company to share in the risk from customers using what it makes).

[290] *See* Charlotte A. Tschider, *Medical Device Artificial Intelligence: The New Tort Frontier*, 46 B.Y.U. L. REV. 1551, 1595 (2021) ("[S]tart-ups most frequently will spend their capital on development and proofs of concept, rather than compliance measures. The goal is to create something that works, rather than to expend capital on proving safety for a large population."). S*ee generally* Daron Acemoglu, *Harms of AI* (Nat'l Bureau of Econ. Rsch., Working Paper No. 29247, 2021), https://www.nber.org/papers/w29247 [https://perma.cc/YZY6-487S] (© 2021 by Daron Acemoglu. All rights reserved.) (detailing potential issues from increasing AI use in society).

[291] *See* Weiser, *supra* note 205, at 2033 n.124; Bryan H. Choi, *Institutional Choice for Software Safety Standards*, 73 HASTINGS L.J. 1461, 1471 (2022).

[292] *See* Weiser, *supra* note 205, at 2033–34; Darrell Rigby, Jeff Southerland & Hirotaka Takeuchi, *Embracing Agile*, HARVARD BUS. REV. (May 2016), https://hbr.org/2016/05/embracing-agile [https://perma.cc/TE3W-VR7M] (describing different approaches to agile techniques).

[293] Tschider, *supra* note 290, at 1605 ("FDA review is incompatible with the realities of AI because it is a linear process, designed for product development lifecycles."); *see also* Matthew Gaske, *Artificial Intelligence Regulation, Minimum Viable Products, and Partitive Innovation*, 73 EMORY L.J. ONLINE 17 (2023).

[294] *See* Tschider, *supra* note 290, at 1590–92; Fletcher, *supra* note 148 at 300 ("When algorithms cause market disruptions or distortions, it is necessary to identify which legal person ought to be held responsible. But this inquiry is not as straightforward as it initially appears."); Lior, *supra* note 102, at 1109 n.25 (noting that AI systems are legally not capable of being liable under the law).

[295] Guihot et al., *supra* note 5, at 418; Cofone, *supra* note 182, at 190–91.

as a doctrine arose to confront new technologies.[296] Products liability generally requires accountability from "the seller or manufacturer of a defective product in a condition that is unreasonably dangerous."[297] A major question for foundation models under this approach is where to draw the line of reasonability in a profoundly powerful, multiuse model that users are rapidly adopting across different contexts to reap competitive advantages.

Another well-documented issue with AI and tort liability is that the unexpected and convoluted nature of advanced AI systems renders a plaintiff establishing causation an ambitious prospect.[298] One approach to remedying this complication is lawmakers identifying constructive or actual knowledge of the potential for harm as the necessary mental state. Such an approach is already analogous to the standard necessary to violate the Bankruptcy Code's automatic stay injunction protecting bankruptcy estate property when a case commences.[299] Specifically, this standard separates the mental state of intending the consequences of an action from the mental state to mean to do an action that ends up having the impact of violating law.[300]

A different option advocates for removing scienter requirements altogether in favor of "[a] harm-focused framework" that places more emphasis on the result of an AI system's interactions with people than on earlier developer or AI platform rationales for those results.[301] This perspective arguably improves deterrence when compared to intent-focused approaches by circumventing issues with algorithm "explainability."[302] Conversely, these traits also may impair AI developers' defenses that depend on users' understanding and appreciation of the potential risks and shortcomings inherent to foundation models.[303]

[296] Chagal-Feferkorn, *supra* note 227, at 77.

[297] *Id.*

[298] Lior, *supra* note 102, at 1111.

[299] *See generally* Matthew Gaske, *Connecting* Kawaauhau v. Geiger *to 11 U.S.C. § 362(K): Considering the "Willful" Requirement for Automatic Stay Enforcement Through* IRS v. Murphy, 15 N.Y.U. J.L. & BUS. 101 (2018) (discussing the bankruptcy automatic stay's "willful" standard for liability).

[300] Cuffee v. Atl. Bus. & Cmty. Dev. Corp. (In re Atl. Bus. & Cmty. Corp.), 901 F.2d 325, 329, (3d Cir. 1990) (quoting In Re Bloom, 875 F.2d 224, 227 (9th Cir. 1989)) ("A 'willful violation' does not require a specific intent to violate the automatic stay. Rather, the statute provides for damages upon a finding that the defendant knew of the automatic stay and that the defendant's actions which violated the stay were intentional.").

[301] Fletcher, *supra* note 148, at 265–66.

[302] *Id.*; Lior, *supra* note 102, at 1111 ("[A]s of now, AI decisions are opaque, unpredictable, and ultimately inexplicable.").

[303] Amy L. Stein, *Assuming the Risks of Artificial Intelligence*, 102 B.U. L. REV. 979, 998–1008 (2022).

In certain cases where innovation is particularly desirable as a policy matter, it might be more appropriate for ex-post active supervision to be a key incentive system for foundation AI models that are subject to emergent properties.[304] Under ex-post active supervision, registration proposals would be helpful for regulators collecting information to better understand how this trait relates to existing law. This would also shift some of the compliance burden onto AI system developers.[305]

Another proposal is analogous to the statements executives make in financial statements that place their imprimatur on the contents of those reports and accept legal consequences for serious flaws.[306] This approach facilitates regulators' finding of liability and addresses many foreseeable risks explicitly covered in such reports, but it does not address the definitionally unforeseeable events from models' emergent properties.[307] Another way to impute liability is through respondeat superior, where the user or designer of the AI is the principal and the AI system is the agent.[308] Notably, even within this targeted-liability regime, second-order tradeoffs persist. Specifically, "[i]ndividual responsibility could lead to decreased diligence in monitoring fellow key function holders."[309] However, this may consolidate additional market power within large-scale AI platform providers that can best police their industry.[310] In any event, an individual or group should take end responsibility for the AI system's output.[311]

[304] Cofone, *supra* note 182, at 191.

[305] Gregory Scopino, *Preparing Financial Regulation for the Second Machine Age: The Need for Oversight of Digital Intermediaries in the Futures Markets*, 2015 COLUM. BUS. L. REV., 439, 499–500; Scherer, *supra* note 255, at 379–80, 84–85; *see also* Fletcher, *supra* note 148, at 305 (describing a similar requirement in fintech: "persons responsible for design, development, or modification of an algorithmic trading program must be registered as a 'Securities Trader' with the Financial Industry Regulatory Authority and pass a qualifying exam").

[306] Fletcher, *supra* note 148, at 323.

[307] *See* Ryan Calo, *Robotics and the Lessons of Cyberlaw*, 103 CALIF. L. REV. 513, 542–45 (2015); *see also* Jason Wei, *137 Emergent Abilities of Large Language Models*, JASON WEI, https://www.jasonwei.net/blog/emergence [https://perma.cc/JT6B-Y77P] (last visited Sept. 30, 2023) (listing observed instances of emergent abilities in foundation models).

[308] Lior, *supra* note 102, at 1164; *see also generally* Lior, *supra* note 181. *But see* W. Robert Thomas, *Corporate Criminal Law Is Too Broad—Worse, It's Too Narrow*, 53 ARIZ. STATE L.J. 199, 231–39 (2021) (criticizing the logic of the respondeat superior doctrine).

[309] Buckley et al., *supra* note 135, at 77.

[310] *See id.*

[311] Lior *supra* note 102, at 1164; *see also* Magnuson, *supra* note 5, at 363 (describing need for people to rely on their own expertise as a check on AI systems).

Proposals also support a strict liability regime for AI harms.[312] One ethical rationale is that AI use creates risks for people who have not, in turn, created the same kind of risks presented by AI for those people and entities that create or provide AI services.[313] In each case though, the preliminary cause of the AI-imposed harm may mitigate a strict liability approach's value, such as future benefits from building on the users' experiences with past AI models.[314] This can also occur when penalizing the scrutinized party seems inequitable, such as when the harm results from intervening forces outside of the party's reasonable control.[315] Nevertheless, strict scrutiny may be a suitable option because, *inter alia*, negligence's emphasis on "reasonability" may eventually force a standard based not on the typical reasonably prudent person standard, but rather on an AI's superior performance on the task at issue.[316] Moreover, this regime reportedly better accounts for the difficulty in assessing algorithms and acknowledges that "distributive" considerations are satisfied because technology companies are the foreseeable payors.[317] Another benefit of this system is that it considers large-scale risk from an appropriately high-level perspective, forcing policymaking decisions of social preferences to make legitimate choices on the potential trade-offs with AI benefits and harms.[318]

However, strict liability standards may be inadequate because they underestimate business teams' creativity.[319] Additionally, a strict liability regime may impair a developer's actionable foreseeability and act as a protecting moat for AI-market incumbents.[320] But there are ways across the water. The software and AI space has a strong open-

[312] *See, e.g.*, Anat Lior, *AI Strict Liability Vis-à-Vis AI Monopolization*, 22 COLUM. SCI. & TECH. L. REV. 90, 94–97 (2020).

[313] *See* Lior, *supra* note 102, at 1114–15 (applying a "[n]onreciprocal [p]aradigm" towards allocation of AI-use risk).

[314] *Id.* at 1129.

[315] *Id.* at 1129–30.

[316] *Id.* at 1119 (at 1120 also considering "fairness and justice" motivations).

[317] *Id.* at 1124–26.

[318] *Id.* at 1128, 51 ("The nonreciprocal approach is a tool that enables us to make policy decisions with regard to what activities are too dangerous for us as a society to endure and what activities are acceptable, welcomed, or even encouraged despite the risks they will certainly inflict.").

[319] Mihailis Diamantis, *Employed Algorithms: A Labor Model of Corporate Liability*, 72 DUKE L.J. 797, 850 (2023); *see also* Chagal-Feferkorn, *supra* note 227, at 93 (identifying the categories of product producers can do to increase safety as "the size of the matrix of parameters the algorithm must consider before making a decision, the dynamic nature of the relevant professional knowledge, the lack of clear right choices, and the extent of trade-off between safety and efficiency" (footnote omitted)).

[320] Lior, *supra* note 102, at 1164; *see also* Magnuson, *supra* note 5, at 358 (describing anticompetitive effects of abundant proprietary information).

source ethos and robust resources to that end.[321] A material level of data, which can vary based on a model's purpose, is necessary for training accurate models and identifying underlying trends in that data.[322] Entrepreneurs and smaller companies can use open-source datasets and open-source web crawlers to help train their models.[323] This counterargument assumes, however, that startups or individuals can both obtain these datasets and tools and employ them legally. This may not be the case.[324] Yet a key, and generally unanswered, question underlying these discussions is how drastically does a developer need to alter a preexisting program or system to avoid liability when the fundamental workings of the program are obscure.[325]

While these smaller companies with fewer resources may benefit from the improved legal clarity of simple rules with predictable outcomes, strict liability standards may contribute to an ossification of the status quo in the technology industry. The lack of a limiting principle, especially in the context of a minimum-viable-product development, exacerbates this issue. This is particularly true when technology incumbents can leverage data network effects and aggressively acquire AI startup companies.[326] As an example of the

[321] *See* Matt Bornstein & Rajko Radovanovic, *Supporting the Open Source AI Community*, ANDREESSEN HOROWITZ (Aug. 30, 2023), https://a16z.com/supporting-the-open-source-ai-community/ [https://perma.cc/4MPM-QSJL]; Alex Engler, *How Open-Source Software Shapes AI Policy*, BROOKINGS INSTITUTION (Aug. 10, 2021), https://www.brookings.edu/articles/how-open-source-software-shapes-ai-policy/ [https://perma.cc/WJV8-VLDW].

[322] Magnuson, *supra* note 5, at 360–61 (explaining the issues accompanying "low-prevalence data").

[323] Peter Wayner, *22 Open Source Datasets to Boost AI Modeling*, VENTUREBEAT (Apr. 7, 2022), https://venturebeat.com/data-infrastructure/22-open-source-datasets-to-fuel-your-next-project/ [https://perma.cc/TR9S-KCKS]; Edmund L. Andrews, *The Open-Source Movement Comes to Medical Datasets*, STAN. U. HUMAN CENTERED INTEL. (Aug. 2, 2021), https://hai.stanford.edu/news/open-source-movement-comes-medical-datasets [https://perma.cc/24LX-ZSKQ]; Cam Dilmegani, *In-Depth Guide to Top 15 Open Source Web Crawlers in 2023*, AI MULTIPLE (Mar. 6, 2023), https://research.aimultiple.com/open-source-web-crawler/ [https://perma.cc/8F3D-2WSU]; *see also What is a web crawler?: How Web Spiders Work*, CLOUDFLARE, https://www.cloudflare.com/learning/bots/what-is-a-web-crawler/ [https://perma.cc/3QWF-VDM3] (last visted Mar. 7, 2023) (describing this kind of computer program).

[324] *See* Van Loo, *supra* note 11, at 836, 72 (describing auto rental companies' use of, *inter alia*, the Computer Fraud and Abuse Act to prevent data scraping of their websites and additionally articulating that "Sellers' current use of the law to freeze the flow of readily available data runs counter to the longstanding consensus in law and economics that good legal interventions should generally remove market information asymmetries, not create them"). *See generally* Alex Luscombe, Kevin Dick & Kevin Walby, *Algorithmic Thinking in the Public Interest: Navigating Technical, Legal, and Ethical Hurdles to Web Scraping in the Social Sciences*, 56 QUALITY & QUANTITY 1023, 1035 (2022).

[325] *See* Lior, *supra* note 181, at 1057–60.

[326] Lior, *supra* note 102, at 1134–38; Van Loo, *supra* note 11, at 828–29 (discussing AI network effects and incumbents' purchase of newer companies generally).

former, some criticism warns that model users might inadvertently be contributing intellectual property through an AI provider's tracking of platforms usage; these users therefore would not have a benefit to bargain for appropriate compensation.[327] To illustrate the latter, Apple, Alphabet, and Microsoft have all acquired a substantial number of less-mature AI companies after 2010.[328]

Professor Matthew Scherer presents an approach that considers ways to incentivize review of AI systems in conjunction with governmental information collection.[329] He proposes a process where companies can receive "limited tort liability" if a government office reviews their AI products and determines that the products are "certified as safe"; otherwise, the products are subject to strict liability.[330] Additionally, that government office would backstop liability from insolvent AI developers.[331] Foundation models complicate this proposal because their multifaced abilities raise the question of which agency is best suited to understand whether those models use cases are safe or whether conflicting messaging from regulators would impair consumer confidence in the approach. Additionally, the potential liability from foundation models could well exceed an amount the government would be willing to provide for the insolvent backstop.[332]

Nor is algorithmic explainability a coverall solution for regulatory enforcement and governmental comprehension of regulated activities; rather, it may reflect a performance trade-off.[333] Moreover,

[327] *See* Eric Sunray, *Train in Vain: A Theoretical Assessment of Intermediate Copying and Fair Use in Machine AI Music Generator Training*, 13 AM. U. INTELL. PROP. BRIEF 1, 8 (2021); *see also* Catrina Denvir et al., *The Devil in the Detail: Mitigating the Constitutional & Rule of Law Risks Associated with the Use of Artificial Intelligence in the Legal Domain*, 47 FLA. ST. U.L. REV. 87–88 (2019) ("ML systems learn from an individual's interaction with the system and where a cloud-based solution is used, an individual is interacting with the software provider's systems and servers. Not all welcome the fact that software developers benefit from the expertise of professional users."); *cf.* Gerrit De Vynck, *AI Learned From Their Work. Now They Want Compensation.*, WASH. POST (July 16, 2023), https://www.washingtonpost.com/technology/2023/07/16/ai-programs-training-lawsuits-fair-use/ [https://perma.cc/6HHQ-BJ7B] (discussing the analogous situation of people seeking compensation for AI company use of content posted online).

[328] *The Race For AI: Here Are The Tech Giants Rushing To Snap Up Artificial Intelligence Startups*, CB INSIGHTS (Sept. 17, 2019), https://www.cbinsights.com/research/top-acquirers-ai-startups-ma-timeline/ [https://perma.cc/756Q-9UE7].

[329] Scherer, *supra* note 255, at 379–80, 84–85.

[330] *Id.* at 394–95.

[331] *Id.* at 394.

[332] *See* Tom Johansmeyer, *Cybersecurity Insurance Has a Big Problem*, HARVARD BUS. REV. (Jan. 11, 2021), https://hbr.org/2021/01/cybersecurity-insurance-has-a-big-problem [https://perma.cc/4CTK-22H8].

[333] Fletcher, *supra* note 148, at 315 ("Algorithms that are built to be explained are less complex than those that are black boxes. Notably, the reduced complexity that increases the algorithm's transparency and explainability also decreases its reliability.").

one's inability to track and understand the decision-making process for deep network learning models becomes exacerbated as the system operates, absorbing more data and adjusting its processes as it learns.[334] Additionally, constructing a blanket requirement for explainability AI imposes substantial hurdles for AI developers, diminishes the risk-reward trade-off of creating new technologies, forces companies to choose between the benefits of AI systems and protecting their intellectual property assets, and taxes regulator ability to actually assess these algorithms.[335] In the foundation model context, this trade-off—and the above-mentioned liability potential—might apply mainly to less-stringent liability regimes. Some foundation models may fall outside the purview of this tradeoff because of their potential for emergent properties, which can be a roadblock to the specific foreseeability of a model's output and effects.[336]

2. Insurance

Briefly, another risk-shifting mechanism attracting attention is insurance coverage for AI-related damage.[337] In the analogous financial context, self-insurance and FDIC protection of consumer accounts are commonplace.[338] A good proxy for understanding the dynamics of this insurance is to look at cybersecurity, one of AI's component risks.[339] However, the market for insurance protection against cyberattacks, for example, seems to be at a relatively nascent stage.[340]

For a more top-down approach, governments could create "a general 'social robotic insurance'" where governments "would be

[334] Christopher K. Odinet, *Securitizing Digital Debts*, 52 ARIZ. ST. L.J. 477, 513–14 (2020).

[335] Fletcher, *supra* note 148, at 316.

[336] *See* Cofone, *supra* note 182, at 184–85 (distinguishing between predictability of elemental AI models versus sophisticated AI models).

[337] *See, e.g.*, Lior, *supra* note 102, at 1158 (describing the potential of network theory to "help . . . insurance companies quantify the damage" caused by AI systems); *see also* Daniel J. Gervais, *Towards an Effective Transnational Regulation of AI*, 38 AI & SOCIETY 391, 400 (2023). *See generally* Anat Lior, *Insuring AI: The Role of Insurance in Artificial Intelligence Regulation*, 35 HARV. J. LAW & TECH. 467 (2022) (providing helpful discussion of the tradeoffs of AI insurance and elaborating on insurance's potential influence on AI development and use).

[338] Lin, *supra* note 75, at 617–18; Van Loo *supra* note 11, at 876.

[339] *See supra* Section III.A.2.

[340] *See* U.S. GOV'T ACCOUNTABILITY OFF., GAO-21-477, CYBER INSURANCE: INSURERS AND POLICYHOLDERS FACE CHALLENGES IN AN EVOLVING MARKET 6, 10, 13, (May 20, 2021), https://www.gao.gov/assets/gao-21-477.pdf [https://perma.cc/X3S5-RFD9] (noting increased premiums, number of insured parties, and a lack of past information to generate payout expectations). *See generally* H. Bryan Cunningham & Shauhin A. Talesh, *Uncle Sam Re: Improving Cyber Hygiene and Increasing Confidence in the Cyber Insurance Ecosystem Via Government Backstopping*, 28 CONN. INS. L.J. 1, 52 (2021) (offering legislative text for a bill "[t]o ensure the continued financial capacity of insurers to provide coverage for risks from cyberattack").

subsidizing innovation by subsidizing insurance premiums."[341] This route would explicitly externalize some of the AI industry's costs, which would benefit innovative new entrants and support industry growth—a considerable benefit in the competitive context.[342] Alternatively, critics view this tax as a more targeted and politically legitimized way to formalize, measure, and reapportion a de facto subsidy that AI companies receive.[343] Specifically, this subsidy exists as the current "algorithmic accountability gap" where "victims of corporations . . . are more likely to be left footing the bill for injuries that algorithms cause."[344]

This approach might be helpful doctrinally because a subtle but key distinction is the relationship between an employer and an AI model, which can be dispositive as to insurance coverage. Specifically, between "a human principal and an AI agent, the latter cannot be found liable, so technically the principal is not vicariously liable but rather primarily liable."[345] Additionally, insurance might be a better vehicle for confronting the pacing problem because new policies and practices can change quicker than regulation.[346]

Conversely, an insurance approach may be more fundamentally difficult because of the lack of information with respect to potential financial fallout from harm and how it would occur.[347] Also, "[m]oral hazard," which describes a party's risk-seeking behavior when others bear the downside risk,[348] may occur with insurance inadequately hedged by contracting.[349] This is particularly alarming with foundation models, as their multi-utility can greatly compound the harmful effects of even a small oversight. Accordingly, these problems might not illustrate risk, but rather "uncertainty," which instead is "where the likelihood of the peril is nonquantifiable."[350] Such irreducibility can

[341] Cofone, *supra* note 182, at 190 n.111 (asserting the existence of such support for defense); *see also Lior*, *supra* note 337, at 493–95; *cf. Superfund*, U.S. ENV'T PROT. AGENCY, https://www.epa.gov/superfund [https://perma.cc/ES66-3UXY].

[342] *Id.*

[343] Diamantis, *supra* note 319, at 826.

[344] *Id.*

[345] Lior, *supra* note 181, at 1099–1100.

[346] Lior, *supra* note 337, at 484–85.

[347] *Id.*

[348] Will Kenton, *Moral Hazard: Meaning, Examples, and How to Manage*, INVESTOPEDIA (June 6, 2023), https://www.investopedia.com/terms/m/moralhazard.asp [https://perma.cc/GTN5-WD44].

[349] Lior, *supra* note 337, at 499 ("Insurance inherently removes, or at the very least reduces, insureds' incentives to prevent harm, since they know that they will not suffer liability as a consequence.").

[350] Daniel A. Farber, *Uncertainty*, 99 GEO. L.J. 901, 903 (2011).

complicate a policymaker's use of insurance as a solution to AI's unintended effects.[351]

Regardless of the pressure point that regulators lean on to solve these public policy concerns, there will be tradeoffs that warrant a focus on executive and board of director duties with respect to AI use in business.[352] Moreover, the Innovation Trilemma suggests that this circumstance will mean more complex rulemaking that may cause confusion and extra costs.[353] Compared to the current regulatory environment for foundation models in the United States, that is likely a sound observation.[354] However, methods do exist to, in aggregate, regulate and promote public benefit, even with reduced legacy complexity.[355] This responsibility escalates as technological capabilities improve.

C. Mitigating Legal Murkiness Through Foundation Model Harm Reduction

The abstracted Innovation Trilemma indicates that foundation models are such a critical resource to countries and companies that regulatory regimes will likely prioritize innovation and mitigation of systemic risk over legal clarity. But low clarity does not necessarily indicate there cannot be meaningful governmental influence and oversight of a quickly evolving industry.[356] While research projects seek a full accounting of the various proposals thought to best balance innovation and social welfare,[357] the original Trilemma suggests that an opaque legal regime could still promote innovation and mitigate systemic risk.[358] Ultimately, the best way to mitigate risk or harm may be to attach regulatory burdens to the sliding scale of innovation itself.

[351] *Cf.* S. Nuri Erbas & Chera L. Sayers, *Institutional Quality, Knightian Uncertainty, and Insurability*, at 4 INT'L MONETARY FUND (2006).

[352] Buckley et al., *supra* note 135, at 77–78.

[353] Brummer & Yadav, *supra* note 29, at 242.

[354] *See, e.g.*, Ashley Gold, *AI Rockets Ahead in Vacuum of U.S. Regulation*, AXIOS (Jan. 30, 2023), https://www.axios.com/2023/01/30/ai-chatgpt-regulation-laws [https://perma.cc/L5GT-FCSG].

[355] *See* Peter Cihon, Jonas Schuett & Seth D. Baum, *Corporate Governance of Artificial Intelligence in the Public Interest*, INFO. 13–17 (2021), https://www.mdpi.com/2078-2489/12/7/275 [https://perma.cc/CT53-HCRR].

[356] *Cf.*, Casey & Lemley, *supra* note 7, at 341 (concluding in the face of a definitional impossibility, "mitigation is our only hope").

[357] *See, e.g.*, Budish, *supra* note 206, at 281 ("[A]cross more qualitative risk governance frameworks, we see the acknowledgement that risk governance is less a sterile scientific endeavor and more a messy political one.").

[358] *See* Brummer & Yadav, *supra* note 29, at 297–306.

1. Collaboration

Though not a novel observation,[359] collaboration among governmental entities and with private industry will be essential in a hyper-detailed regulatory regime to preemptively understand public policy concerns and efficiently guide innovation away from projects that may result in costly legal scrutiny and uncertainty. Through co-creation and communication, regulation may be better poised to enable new approaches to innovation.[360] This allows foundation models' competitive context to counterbalance "bureaucratic inertia" that stalls aggressive innovation and enables more innovation-accommodating regulatory structures to assist the AI industry. In this fast-moving, technical field, a typical regulatory approach courts "disastrous results."[361]

A preliminary question then is how agencies should collaborate with one another.[362] Greater government investment into macro-coordination would be helpful here. One option for this coordination role is the Office of Information and Regulatory Affairs, which has "review[ed] agency regulatory actions for consistency with presidential priorities, statutory mandates, and, notably, other agencies' rules."[363] Another option that could be legislatively empowered is the Networking & Information Technology R&D Program, which is "the U.S. federal government's primary coordinating body for federal R&D in advanced digital technologies" and already has "25 member agencies and more than 60 participating agencies."[364] Additional options exist, but these examples simply illustrate that a new federal agency is not an absolute

[359] *Id.* at 282–302.

[360] Joshua Sarnoff, *The Likely Mismatch Between Federal Research & Development Funding and Desired Innovation*, 18 VAND. J. ENT. & TECH. L. 363–93 (2016) (describing the debate and potential context dependence of the "Porter Hypothesis" proposing "that firms which respond to stringent regulation by developing new technologies have a 'first mover" advantage and can capture the market for their products/services").

[361] Weiser, *supra* note 205, at 2056–57.

[362] *See* Brummer & Yadav, *supra* note 29, at 297–301 (discussing how agencies may work together to regulation fintech innovation).

[363] Freeman & Rossi, *supra* note 265, at 1179; *see* Aram A. Gavoor & Raffi Teperdjian, *A Structural Solution to Mitigating Artificial Intelligence Bias in Administrative Agencies*, 89 GEO. WASH. L. REV. ARGUENDO 71, 75–77 (2021). *See generally* Cass R. Sunstein, *The Office of Information and Regulatory Affairs: Myths and Realities*, 126 HARV. L. REV. 1838 (2013) (describing the Office's operation from the perspective of that Office's former head). *But see* Solow-Niederman, *supra* note 17, at 674 ("[W]e presently lack the requisite preconditions for collaborative AI governance."). *See generally* Nina A. Mendelson & Jonathan B. Wiener, *Responding to Agency Avoidance of OIRA*, 37 HARV. J.L. & PUB. POL'Y 447 (2014) (addressing potential agency desires and strategies for circumventing that Office's review).

[364] Zhang et al., *supra* note 255.

necessity in all cases, though whichever existing entity receives this responsibility would likely need additional resources.

Concerning initial regulatory coverage for a novel technological field, an agency or other governmental body may consider more informal communications as a malleable approach to help guide company behavior as a field develops.[365] To the extent formal regulation becomes necessary, such regulation should consist of standards that adopt "a principle-based approach," which allows leeway for governmental entities to promulgate regulations guiding AI market developments.[366] Indeed, the way past "the regulatory dichotomy of recklessness or paralysis is a willingness to move beyond the expectation of finality that surrounds regulatory decision-making."[367]

A series of mechanisms can aid innovation with relatively low risk to the public. First, "regulatory sandboxes" is an approach well suited for the foundation model context, assuming the model network does not extend outside the parameters of that sandbox.[368] Within a sandbox, company participants are essentially given exemptions from particular regulations to experiment with a product on a small set of customers who are aware of the experimental nature of that product.[369] Notably, anticompetitive concerns exist for allowing large, established technology companies to participate in policy sandboxes.[370] These concerns may motivate the redistribution of systemic risk and encourage companies to develop in a manner preempting the algorithmic monoculture described above.[371] Next, innovation hubs offer a central pathway for companies to ask a regulator for informal

[365] Moran Ofir & Ido Sadeh, *More of the Same or Real Transformation: Does Fintech Warrant New Regulations?*, 21 HOUS. BUS. & TAX L.J. 280, 313 (2021); *see also* Brummer & Yadav, *supra* note 29, at 275–78, 283. *But see* Gao & McDonald, *supra* note 285, at 949, 951 (explaining that "ventures that incorporate regulatory considerations into their initial strategy formulation have more trouble gaining market traction than those that delay doing so" and recognizing the informational benefits a company can gain by resisting "regulatory pressures").

[366] Fenwick, *supra* note 66, at 590.

[367] *Id.*

[368] Brummer & Yadav, *supra* note 29, at 283. *See generally* Hilary J. Allen, *Regulatory Sandboxes*, 87 GEO. WASH. L. REV. 579 (2019).

[369] Fenwick, *supra* note 66, at 591–93; *see also* Brummer & Yadav, *supra* note 29, at 292 ("[R]ather than be subject to restrictive or complex rules that elevate regulatory risk and potentially stifle innovation, the sandbox offers a means of testing new ideas in a simplified, interactive regulatory environment.").

[370] *See* Savannah P. Schaefer, *Save Our Sandbox: A Prospective Approach to Big Player Participation in the Unlicensed Spectrum Sphere*, 15.1 COLO. TECH. L.J. 233, 240–47 (2016) (discussing competitive issues in the analogous context of "unlicensed spectrum bands").

[371] Ofir & Sadeh, *supra* note 365, at 315–17.

feedback on potential regulatory requirements.[372] In the European Union for example, this approach has allowed businesses to familiarize themselves with technological tools in a low-risk and low-cost setting.[373]

Globally, governments may also need to coordinate the technical aspects of foundation models. One such proposal entails public entities first assisting newer AI companies with their models through the cultivation and distribution of "fair and equitable training dataset[s]"[374] because of their routine yet expansive data-collection and organization activities.[375] In other cases, when regulators try to assess technical code, a regulating body should either internally collaborate on those assessments or collaborate with other agencies in overlapping jurisdictional space to review each other's work.[376]

Within this model, governmental entities might work best as an information clearinghouse for relatively safe approaches within an industry, and those suggestions would be persuasive because of their potential to protect companies from liability.[377] In particular, systemic barriers impair widespread understanding of how or why an AI model went awry.[378] These can be problematic in the foundation-AI context because a characteristic of "networked innovation systems" is that "small-size errors do not necessarily result in small-scale problems."[379]

[372] Petja Ivanova, *Cross-Border Regulation and Fintech: Are Transnational Cooperation Agreements the Right Way to Go?*, 24 UNIF. L. REV. 367, 389 (2019).

[373] Maurits Butter, Govert Gijsbers, Arjen Goetheer & Kristina Karanikolova, *Digital Innovation Hubs and Their Position in the European, National and Regional Innovation Ecosystems*, *in* REDESIGNING ORGANIZATIONS: CONCEPTS FOR THE CONNECTED SOCIETY 46–47 (Denise Felder ed., 2020).

[374] Laura Weidinger, John Mellor, Maribeth Rauh, Conor Griffin, Jonathan Uesato, Po-Sen Huang, Myra Cheng, Mia Glaese, Borja Balle, Atoosa Kasirzadeh, Zac Kenton, Sasha Brown, Will Hawkins, Tom Stepleton, Courtney Biles, Abeba Birhane, Julia Haas, Laura Rimell, Lisa Anne Hendricks, William Isaac, Sean Legassick, Geoffrey Irving & Iason Gabriel, *Ethical and Social Risks of Harm from Language Models*, DEEPMIND 12 (Dec. 8, 2021), https://arxiv.org/pdf/2112.04359.pdf [https://perma.cc/9XLJ-JQTY].

[375] Ricard Munné, *Big Data in the Public Sector*, *in* NEW HORIZONS FOR A DATA-DRIVEN ECONOMY, 195, 199 (Jose Maria Cavanillas, Edward Curry & Wolfgang Wahlster eds., 2016).

[376] J.B. Ruhl & James Salzman, *In Defense of Regulatory Peer Review*, 84 WASH. U. L. REV. 1, 61 (2006) (cautioning, however, that it "may well prove unwise to mandate peer review across the board for agency actions . . . without a clear understanding of the real extent of the problem peer review is supposed to address").

[377] Weiser, *supra* note 205, at 2023 n.61.

[378] *See* Shur-Ofry, *supra* note 125, at 361, 372 ("Much like positive information goods, negative information is easy to duplicate. Yet, unlike positive knowledge, negative knowledge does not easily translate into material objects, and it is extremely difficult to trace and control its unauthorized use by third parties.").

[379] *Id.* at 368.

2. Strategic Pauses

Sometimes an AI system needs calibration time to examine itself or its surroundings for a better understanding of what to do during a process.[380] Likewise, monitors require time to assess a situation when a process malfunctions.[381] Returning to the financial context, strategic pauses occur when financial markets' "circuit breakers" stop trading because volatility is too great.[382] In fact, Microsoft has endorsed "a requirement that systems used in critical infrastructure can be fully turned off or slowed down."[383] Somewhat ironically, however, system monitors would require programming to prevent algorithmic systemic issues, especially as machines begin to outpace humans in conducting certain operations.[384] Strategic pauses are mainly useful if the catalyzing occurrence is detected with time to react before harm occurs, which requires AI companies and those systems' users—as well as potentially regulators—to track and analyze data in real time.[385]

Nassim Nicholas Taleb's "barbell" concept is useful here.[386] This is "a bimodal strategy," which heavily pursues exposure to both (1) the benefits of high-risk activities (e.g., the foundation model's powerful but unpredictable unexpected operations), and (2) the benefits of very low-risk activities (e.g., the foundation model pausing to conducting self-

[380] *See generally* Josh Bongard, *Resilient Machines Through Continuous Self-Modeling*, 314 SCI. 1118 (2006) (describing a robot's process of understanding its own form).

[381] Chuck Mikolajczak, *Explainer: Wall Street's Market Glitches and the Repercussions*, REUTERS (Jan. 25, 2023), https://www.reuters.com/markets/us/wall-streets-market-glitches-repercussions-2023-01-24/ [https://perma.cc/FU2M-M3ZJ] (describing trading halts on exchanges); *see* Paul Ohm & Jonathan Frankle, *Desirable Inefficiency*, 70 FLA. L. REV. 777, 782 (2018) ("[I]nefficiency has been intentionally injected into systems, dialing back the raw speed and power our information-age conditioning instinctually reveres—all as a means for promoting nonefficiency values.").

[382] Lin, *supra* note 182, at 604; *see also* Marchant & Stevens, *supra* note 140, at 244 (characterizing "kill switches" as a form of "resilience by design"). *See generally* Andy Kessler, *AI Simply Needs a Kill Switch*, WALL ST. J. (Apr. 23, 2023), https://www.wsj.com/articles/ai-simply-needs-a-kill-switch-regulations-biden-commerce-innovation-musk-pause-spacex-767ae62a [https://perma.cc/7KZR-JE7J].

[383] David McCabe, *Microsoft Calls for A.I. Rules to Minimize the Technology's Risks*, N.Y. TIMES (May 25, 2023), https://www.nytimes.com/2023/05/25/technology/microsoft-ai-rules-regulation.html [https://perma.cc/M77V-VZ6M].

[384] *See* Fletcher, *supra* note 148, at 261 ("Computers running algorithmic trading programs have taken over as the primary 'traders' in the market, while humans execute merely ten percent of all trades today.").

[385] *See* David Orozco, *Compliance By Fire Alarm: Regulatory Oversight Through Information Feedback Loops*, 46 IOWA J. CORP. L. 97, 140; *see also* Fenwick et al., *supra* note 66, at 585 (proposing that regulators get an ex ante view of where regulatory attention may be needed by keeping track of venture investments or other inflows to advancing technology because "start-up companies usually challenge existing rules, laws, and regulations").

[386] TALEB, *supra* note 48, at 159.

assessment or with people's assistance) to account for unknowable and unpredictable elements.[387] In fact, technology luminaries have suggested a macro-strategic pause of six months would be appropriate for government and industry to gather standards and work collectively to provide safer foundation models.[388]

However, pauses to slow or stop AI systems, like those in securities trading, might have a limited ameliorative effect because of the variety of different alternative venues and intermediaries for information flow,[389] which make rapid cutoffs difficult to manage.[390] While this removes a systemic bottleneck risk at a singular and key juncture, this arrangement likely also creates an asymmetric responsibility for a few regulators to monitor multifaceted channels of informational flow in a variety of submarkets comprised, in turn, of a potentially vast array of smaller entities.[391] In sum, strategic pauses are a relatively static way of trying to balance risk and innovation in low-clarity regimes. The next subsection discusses another option: making regulation a sliding scale driven by innovation itself.

3. Linking Harm Avoidance to Pace of Innovation

Foundation models will operate differently than developers and users expect, and downstream processes that depend on the AI system will consequently break.[392] A key question for the regulatory priorities emerging from the Trilemma is what happens following this break. Linking ex-post regulatory requirements to the rate of the AI field's advancement may serve the interests among all three categories. The goal is to reduce the harm that occurs—or tolerance of that harm—when AI models act unexpectedly by more than merely reducing the

[387] *Id.* at 161.

[388] *Pause Giant AI Experiments: An Open Letter,* FUTURE OF LIFE INST. (Mar. 22, 2023), https://futureoflife.org/open-letter/pause-giant-ai-experiments/ [https://perma.cc/P6F6-E95U].

[389] Lin, *supra* note 143, at 1298 (noting the "balkanization of the marketplace" and the surge in available trading venues: "In other eras, a failure of the New York Stock Exchange would have brought a majority of equity trading in the United States to a halt . . . In 2016, there were over twenty registered national exchanges and around seventy total trading venues for securities and futures trading.").

[390] Jón Daníelsson, Robert Macrae & Andreas Uthemann, *Artificial Intelligence and Systemic Risk*, 140 J. BANKING & FIN. 3 (2022), https://doi.org/10.1016/j.jbankfin.2021.106290 [https://perma.cc/Q8NF-34WG] (describing permissions in finance as shifting from a physical "rulebook" to "digital logic, allowing programmatic access").

[391] Lin, *supra* note 143, at 1298.

[392] *See* Florian Tramèr, Rohith Kuditipudi & Xuechen Li, *Security and Privacy*, *in* BOMMASANI ET AL., *supra* note 14, at 105–07.

odds of an adverse outcome.[393] Instead, the Trilemma and policymakers' decisions within the Trilemma given the U.S.'s international competitive stance suggest a systemic risk mitigation regime that keeps the pressure on companies to innovate.

First, a system that progressively escalates regulatory requirements as an AI development company matures is one option for balancing risk protection and innovation. These regulatory requirements would link mandated burdens to technical advancements and demonstrated value, which is idiosyncratic to a model's users.[394] Helping new firms enter a concentrated marketplace is a benefit of this approach, which "lower[s] the entry barriers . . . while keeping the sentries at the entry gates."[395] Similarly, regulators can adapt to the modern software development cycle, such as with the FDA's experimentation with periodic review of AI medical products instead of universal application of its typical, arduous ex ante review.[396]

Second, the breadth of foundation models' capabilities, described in Section III.B, renders accurately tailoring regulation to these systems particularly difficult. However, Casey and Lemley suggest institutionalizing practices that allow the law to adapt to changing technological innovation, such as processes to except specific cases from regulation or "sunset clauses" for regulatory policies.[397] They also suggest laws discouraging particular acts rather than trying to classify

[393] *See* J.B. Ruhl, *Complexity Theory as a Paradigm for the Dynamical Law-and-Society System: A Wake-Up Call for Legal Reductionism and the Modern Administrative State*, 45 DUKE L.J. 849, 886 (1996) ("Those systems that have demonstrated sustainability have somehow managed to build into their structures qualities that help them survive the surprises produced by chaos, emergence, and catastrophe."); *see also* Timothy Malloy, *Re-Imagining Risk: The Role of Resilience and Prevention*, 22 NEV. L.J. 145, 186, 193 (2021) (implicitly delineating the concepts).

[394] Dirk Zetzsche, Ross P. Buckley, Janos N. Barberis & Douglas W. Arner, *Regulating A Revolution: From Regulatory Sandboxes to Smart Regulation*, 23 FORDHAM J. CORP. & FIN. L. 31, 97–98 (2017) (proposing four levels of regulation, escalating from "testing," a broader "testing" set in "a regulatory sandbox," a low-impact "licensing" approach for growing, younger companies, and "a full license" for mature companies).

[395] *Id.*

[396] *Statement from FDA Commissioner Scott Gottlieb, M.D. on Steps Toward a New, Tailored Review Framework for Artificial Intelligence-Based Medical Devices*, U.S. FOOD & DRUG ADMIN. (Apr. 2, 2019), https://www.fda.gov/news-events/press-announcements/statement-fda-commissioner-scott-gottlieb-md-steps-toward-new-tailored-review-framework-artificial [https://perma.cc/MMB8-XP5V]; *see also* Tschider, *supra* note 290, at 1604 ("[T]he inscrutability aspect of more advanced AI complicates ongoing and preventative monitoring, or even postmarket surveillance, a crucial part of the FDA regulatory structure that enables the FDA to take action, such as recalling devices to prevent further injury."); W. Nicholson Price II & Arti K. Rai, *Clearing Opacity Through Machine Learning*, 106 IOWA L. REV. 775, 805–06 (2021) (describing an FDA "pilot program"). *See generally* Gaske, *supra* note 293 (discussing the FDA example and considering the relevance of ex ante versus ex post approaches in AI regulation).

[397] Casey & Lemley, *supra* note 7, at 360–61.

technologies or companies.[398] This approach circumvents upstream questions regarding the subjects of regulation in an evolving technical field.[399]

Third, algorithmic opacity is not necessarily fatal to regulation; outcome-focused legal protections, such as disparate impact in the discrimination context, may instead be useful for policing the application of these systems.[400] However, regulators may need to modulate the requisite level of disparate impact, address potential unfamiliarity across regulatory bodies, and determine thresholds for statistical evidence implying causation.[401] Although some argument persists concerning how machine learning can assign weights across multiple inferential layers within a model's architecture, "causal relationships between inputs and outputs may simply not exist, no matter how intuitive such relationships might look on the surface."[402] Thus, outcomes may be more concrete signals for regulators of problems that require their attention in an ever-changing environment.

Innovation in advanced technology that comes to market almost necessarily entails the public shouldering negative externalities.[403] A legal standard focusing regulation on subsequent harm as a function of innovation also enlists the producers and users of foundation models to exercise their cutting-edge expertise (that regulators likely lack) to address problems.[404] Under this legal standard, the emergence of new, negative externalities should be an ongoing concern for regulators and developers alike. Engineering may provide solutions, such as "designing minimally invasive [AI] agents that prefer easily reversible to irreversible actions" or models that weigh a variety of different goals instead of merely one.[405]

[398] *Id.* at 342 (explaining their concept of "Turing's Razor," meaning "whenever possible, establish whether a potential regulated entity is a robot without resorting to explicit, ex ante definitions").

[399] *Id.*

[400] Odinet, *supra* note 334, at 538–41.

[401] *Id.* (emphasis in original); *see also* Coglianese & Lehr, *supra* note 260, at 1191–1205, 1217; *Title VI Legal Manual*, U.S. DEP'T OF JUST., TITLE VI LEGAL MANUAL § VII, https://www.justice.gov/crt/fcs/T6Manual7#M [https://perma.cc/3HSL-GYUU] (last visited Sept. 30, 2023) (describing the relationship between statistics and causation for purposes of disparate impact cases).

[402] Coglianese & Lehr, *supra* note 260, at 1156–57.

[403] Diamantis, *supra* note 319, at 825 ("When corporations limit their liability but not the harmfulness of their conduct, they externalize some of the true costs of their operation. As every economist would predict, this means corporations will use algorithms even when, from a net social welfare standpoint, it would be best if they refrained.").

[404] *See* Casey & Lemley, *supra* note 7, at 343 (recognizing the benefits of technological regulation's focus on end results).

[405] Hendrycks et al., *supra* note 124, at 10.

Customer-specific dependency also arises when a foundation-model provider collects extensive information from individual customers to refine a model for that individual's particular use case. The difficulty of switching to another data provider certainly provides competitive benefits to the provider,[406] but switching also forces the customer to share a provider's risk that the AI system will continue operating as both the provider and customer intend. If an event occurred that disrupts the provider's AI abilities, the technical (as opposed to legal) inability to transfer a customer's data to a customer-accessible repository may cause concern.[407]

However, data portability by itself is not a sufficient buffer to protect consumers.[408] A provider in the onset of crisis might have only limited human resources or technical capabilities to transfer consumer data.[409] Customers who are aware of this breach may exacerbate the problem by causing a "data-bank run," where they rush to obtain copies of their data from a company in turmoil, effectively preventing many customers from successfully porting their data by demand of company resources. A technical solution may exist for providers engaging in data porting, such as the provider creating a data repository on a separate, provider-operated system that encourages consumer confidence through a perception of legitimacy.[410] Another option could perhaps consist of a distinct corporate entity specifically catering to consumers' needs. Because this approach would inhibit a competitive advantage from leveraging data network effects, a regulator or another external legal authority would likely need to impose a solution of this nature.

While the goal of AI regulation is protecting against risk (i.e., the potential for harm), and more importantly, harm itself, AI technology will fail or act unexpectedly at some point, regardless of how

[406] *See* Barak Libai et al., *Brave New World? On AI and the Management of Customer Relationships*, 51 J. OF INTERACTIVE MKTG. 44, 49 (2020).

[407] *Cf.* Brooke Auxier, Lee Raine, Monica Anderson, Andrew Perrin, Madhu Kumar & Erica Turner, *Americans and Privacy: Concerned, Confused and Feeling Lack of Control Over Their Personal Information*, PEW RSCH. CTR. (Nov. 15, 2019), https://www.pewresearch.org/internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information/ [https://perma.cc/N9TU-AVYJ].

[408] *See* Daniel Gill & Wolfgang Kerber, *Data Portability Rights: Limits, Opportunities, and the Need for Going Beyond the Portability of Personal Data* 6 (Oct. 11, 2020) (unpublished manuscript), https://ssrn.com/abstract=3715357 [https://perma.cc/KNZ2-KE74] (explaining that data portability rights might face effectiveness issues due to "transaction costs" and issues of consumer knowledge of their rights).

[409] *See id.*

[410] *Cf.* Russ Kennedy, *Why the Evolution of Data Storage Matters*, FORBES (Mar. 28, 2023, 6:00 AM), https://www.forbes.com/sites/forbestechcouncil/2023/03/28/why-the-evolution-of-data-storage-matters/?sh=1a24b0585809 [https://perma.cc/6PD6-U3TP] (describing the emergence of cloud computing and data retention).

well its creators have attempted to account for risk. Therefore, efforts to reduce the harm that occurs when technology fails are critical and a necessary accompaniment to any AI risk-reduction effort.[411] Foundational models present a particularly difficult case because of their scope; targeted policies to mitigate unforeseen, second-order effects might not be workable.[412]

The relevant question here is how a regulator can increase the costs to malicious or reckless actors while avoiding imposing costs on innovators generally.[413] One approach is regulators pursing a goal of legal clarity in their communication with disparate parties.[414] How can policymakers define different levels of liability when unexpected occurrences characterize AI applications? Indeed, this question supports the use of Professor Gina-Gail Fletcher's "harm-based" approach to regulation, which appropriately accounts for the difficulties in understanding scrutinized parties' state of mind.[415] Professor Fletcher similarly suggests, "if the algorithm's conduct is proven to be harmful to the market, this creates a rebuttable presumption of liability."[416] Instead, she offers the alternative of using "a recklessness standard" based on "whether [the programmer] followed industry norms and standards in designing the algorithm."[417]

Professor Bryan Choi examines a similar idea by looking to the analogous development of automobiles, vehicles that eventually are involved in some kind of accident, and the accompanying "crashworthy doctrine."[418] This theory "holds that a vehicle manufacturer owes a duty to use reasonable care in the design and manufacture of a product to minimize injuries to its users and not to subject its users to an unreasonable risk of injury in the event of a collision or impact."[419] This approach seems to appropriately balance an AI provider's competitive need for innovation with an escalating obligation owed to users as

[411] Bryan Choi, *Crashworthy Code*, 94 WASH. L. REV. 39, 44, 111–13 (2019).

[412] Lin, *supra* note 143, at 1278 (citing J.B. Ruhl & James Salzman, *Mozart and the Red Queen: The Problem of Regulatory Accretion in the Administrative State*, 91 GEO. L.J. 757, 814 (2003)).

[413] *See* Fletcher, *supra* note 148, at 267 (describing deterrence theory).

[414] *Id.* at 272–73; *cf.* Ops. Att'ys Gen., NAT'L ASS'N ATT'YS GEN., https://www.naag.org/issues/civil-law/attorney-general-opinions/ [https://perma.cc/SS2N-NALE] (example of law enforcement entities providing prospective views on specific legal questions).

[415] Fletcher, *supra* note 148, at 266.

[416] *Id.* at 319.

[417] *Id.* at 321.

[418] *See generally* Choi, *supra* note 411.

[419] *Id.* at 45.

technology advances.[420] This approach also disconnects the probability of an occurrence from the impact of that occurrence[421] and may be an adequate driver of both innovation and harmless (or minimally harmful) use.[422] The above modified Trilemma analysis indicates that this tort approach integrates well with the macro-competitive context. Therefore, technological advancement can be accordingly addressed in the AI regulatory space.[423]

V. CONCLUSION

Foundation models present an opportunity for the fundamental advancement of human activities. This opportunity comes with pitfalls that invite public scrutiny and legal restraints. Considering the scope of potential regulatory options to address these issues requires a policymaker, manager, or other stakeholder to assess the context of this technology's use and the evolution of the markets it enables. This Article has used a modified version of the Innovation Trilemma to guide an analysis and understanding of the external circumstances that will continue to inform public policy from a high-level perspective. Policymakers' accepting competition as a central trait of the foundation model ecosystem renders downstream regulatory priorities clearer.

New sources of systemic risk arise from the profound technological change foundation models propagate. These changes unfortunately require more universal attention than may be typical in other industries to achieve regulatory clarity, especially in a relatively consolidated marketplace where model sameness can be problematic. Despite the necessary heightened regulatory attention, the reprioritization of legal clarity does not mean a lack of oversight.

[420] F. Patrick Hubbard, *"Sophisticated Robots": Balancing Liability, Regulation, and Innovation*, 66 FLA. L. REV. 1803, 1853–55 (2014) (describing, in the autonomous vehicle context, the forms of evidence, such as expert testimony, that would be relevant to this kind of approach).

[421] Michal Shur-Ofry & Ofer Tur-Sinai, *Constructive Ambiguity: IP Licenses As a Case Study*, 48 U. MICH. J.L. REFORM 391, 412 (citing TALEB, *supra* note 48, at 234–35) (footnote omitted) ("[R]ather than trying to anticipate and regulate the unpredictable, players in complex systems should concentrate on increasing their capacity to adjust to and even benefit from unforeseen events."); *see also* Nassim Taleb, *"Antifragility" As a Mathematical Idea*, 494 NATURE 430 (2013), https://doi.org/10.1038/494430e [https://perma.cc/WJR7-SVH6].

[422] Keith N. Hylton, *The Law and Economics of Products Liability*, 88 NOTRE DAME L. REV. 2457, 2496 (explaining incentives of the "risk-utility test": "if a manufacturer is considering a design for which the court's likely assessment of the balance between incremental risk and incremental utility is unclear, the manufacturer is likely to have an incentive under the test to err toward safety").

[423] Varun Bhatnagar, *The Evidentiary Implications of Interpreting Black-Box Algorithms*, 20 NW. J. TECH. & INTELL. PROP. 433, 442–47 (2023) (describing the use of the "SHAP" and "Lime" "statistical techniques").

Rather, regulators will need to learn from industry experiences and maintain a flexible disposition. Although such flexibility does not lend itself to predictable future policy, this flexibility allows policymakers and regulators to understand the reality of how foundation models are used.

Of course, counterarguments assume a more skeptical stance regarding the ideas presented in this Article. For example, a massive surge in public-sector spending on government-related advanced AI technologies may mitigate a dependence on private-sector companies for AI development. In turn, this spending may invite a reshuffling of public policy priorities. This Article's purpose nonetheless is to contribute to conversations about how to understand the pressures on foundation model development from a broad perspective. This Article acknowledges the conjoint role of competition, systemic risk, and regulatory comprehensibility for AI regulation outcomes. Recognizing this relationship will hopefully guide a meta-discussion of how policymakers can best prioritize their goals for foundation model regulation with an appreciation of those models' context.