

Lessons from the FDA for AI

[Introduction](#)

[Executive Summary](#)

[Why an Independent Agency? And Why the FDA?](#)

[Methods](#)

[How to Read This Report](#)

[Section 1: The Food and Drug Administration](#)

[Section 2: What Does the FDA Do?](#)

[I. Premarket Approval](#)

[What is it?](#)

[How could it work for AI?](#)

[II. Post-Market Monitoring and Enforcement](#)

[What is it?](#)

[How could it work for AI?](#)

[III. Producing Information and Expertise](#)

[What is it?](#)

[How could it work for AI?](#)

[Section 3: Lessons from the FDA Model](#)

[I. Establishing Efficacy and Safety](#)

[II. From Opacity to Openness](#)

[III. Generating Agency and Rebalancing Power](#)

[Section 4: Challenges for FDA-Style Interventions](#)

[I. Establishing a Regulatory Perimeter](#)

[II. Enabling Robust Enforcement](#)

[III. Navigating the Courts](#)

[1. First Amendment and Content Moderation](#)

[2. Section 230](#)

[3. US Courts and Rulemaking](#)

[IV. Preventing Industry Capture](#)

[Conclusion](#)

[Glossary](#)

[Appendices](#)

[Appendix 1: How AI Is Regulated Today](#)

[Appendix 2: How Does the FDA Work?](#)

[Appendix 3: Comparison of FDA Mechanisms and AI Regulatory Proposals](#)

[Further Reading](#)

[Acknowledgments](#)

Introduction

FDA Introduction

Executive Summary

FDA Executive Summary

Why an Independent Agency? And Why the FDA?

The United States is in a unique moment for AI policy: a groundswell of public interest in artificial intelligence has captured policymaker attention, at a moment when the White House has adopted a historically unique willingness to adopt a more muscular posture toward the tech industry. At the same time, Congress has largely been hamstrung by political squabbling: in an election year, keeping the lights on, let alone passing complex and capital-intensive legislation, remains a hard task. This makes it an opportune moment to step back and directionally align on the best approach to core AI governance questions, orienting ourselves around the world of the possible rather than making incremental tradeoffs in service of the practicable.

Creating a novel AI agency whose primary responsibility would be to regulate the actors responsible for developing and deploying AI systems, either to supplant or to augment existing enforcement mechanisms distributed across the US government, is one of many possible paths forward.¹ The current moment has led to the reinvigoration of what is an old idea: since 2017, a range of stakeholders have proposed such an agency, frequently referencing the construct of an “FDA for AI” alongside proposals for “nutritional labels” for AI systems and other similar approaches to licensing and certification modeled on approaches to food and drug safety (see Box 1).

The FDA is the regulatory agency responsible for ensuring the **safety, efficacy, and security** of the nation’s food, biological, and medical products through a rigorous product-oversight and premarket-approval regime. Such an approach offers the promise of greater regulatory friction that would ensure the burden is on companies to adequately vet their systems for efficacy and potential harm—before, rather than after, public release.² On the other hand, this style of


¹ For a summary of such mechanisms, see Appendix 1.

² See Accountable Tech, AI Now Institute, and EPIC, “Zero Trust AI Governance,” AI Now Institute (blog), August 10, 2023, <https://ainowinstitute.org/publication/zero-trust-ai-governance>; and Gianclaudio Malgieri and Frank Pasquale, “From Transparency to Justification: Toward Ex Ante Accountability for AI,” Brooklyn Law School,

regulatory oversight could enable “check-box certification,” distracting from existing enforcement capabilities and creating opportunities for regulatory capture. If an independent agency is indeed the right path forward for AI regulation, circumventing these challenges—and setting a strong administrative foundation for accountability—would be key.

Given the growing momentum building around models for AI governance, this is a pressing moment to be asking: Is an “FDA for AI” a good idea? And more specifically: By looking deeply into this example, can we become more concrete about what specific regulatory authorities are needed to effectively govern AI?

Box 1: Who is talking about an “FDA for AI”? See:

 RR Copy of FDA Report_Restructure v2

Methods

Rapid Expert Convening

To delve into these questions, the AI Now Institute convened a group of experts who combine decades of experience studying the FDA, the pharmaceutical industry, and artificial intelligence for a rapid deliberation on the following question: Do we need an FDA for AI? The group included former government officials, academic researchers, medical doctors, lawyers, computer scientists, and journalists from a variety of countries for a collective deep dive into lessons from FDA-style regulation and their potential application to the domain of artificial intelligence.

Synthesis of Key Areas of Expert Consensus, and Open Questions

Our conversation was deliberately structured with flexibility to enable the group to explore emergent themes for conversation, and this is reflected in our takeaways: given the state of the conversation, we felt it most appropriate to surface tensions and complexities rather than drive straight to solutions. We also tabled, and took notes on, the many barriers to implementing the governance ideas that surfaced throughout the conversation, constructing a space designed around surfacing the best answers we could come up with rather than what’s practically feasible in the near term.

We published an [interim memo](#) with ten actionable insights from the workshop to represent the rough consensus that emerged and to speak back to the ongoing policy conversation.

The Way Forward

This report goes further: many of the issues surfaced through the workshop require deeper engagement and a more robust reckoning with how the specifics of the FDA model interact with the specifics of the AI ecosystem and current regulatory priorities. As such, we intend this report

to form a foundation for an ongoing conversation about potential paths for AI governance models.

How to Read This Report

This report is divided into four sections. Section 1 outlines the key features of the FDA and briefly summarizes its development since the emergence in the late nineteenth century of the earliest pharmaceutical regulation in the US.

Section 2 provides an overview of the FDA's main regulatory functions for pharmaceuticals, grouping them into three categories. In this section we describe in detail how the FDA discharges these functions and suggest how they could work for AI, surfacing areas of difficulty and complexity.

Section 3 outlines three crosscutting lessons from the experience of the FDA's regulation of pharmaceuticals. Rather than homing in on specific functions like Section 2, it looks at how these functions interact to produce particular outcomes in the sector.

Section 4 discusses four practical challenges for implementing FDA-style interventions in AI: establishing the correct bounds of the AI market; providing AI regulators with the necessary powers to shape industry practices; overcoming legal challenges to the exercise of these powers; and avoiding industry capture.

At the end of the report we offer a glossary of key terms and appendices with information on how AI is regulated today and a detailed mapping of AI regulatory powers against those of the FDA.

Section 1: The Food and Drug Administration

I. The Food and Drug Administration

The Food and Drug Administration (FDA) is a federal agency within the Department of Health and Human Services (HHS), an **executive agency** with a Commissioner and a budget of \$7.2 billion in 2024 (an increase of \$500 million from 2023).³ By comparison, the FTC, the agency responsible for enforcing the consumer protection and competition laws that have dominantly constituted AI regulation in the US, had a total budget of \$430 million in 2023.⁴ The FDA's mandate and scope is broad in nature: it is the primary public health agency responsible for the safety, efficacy, and security of human and veterinary drugs, biological products, and medical

³ Food and Drug Administration, "FY2025 FDA Budget Summary," 2025, <https://www.fda.gov/media/176923/download?attachment>.

⁴ Federal Trade Commission, "Budget and Strategy," June 7, 2013, <https://www.ftc.gov/about-ftc/budget-strategy>.

devices; as well as the food supply, cosmetics, and products that emit radiation.⁵ In this report, we focus more specifically on the oversight process for pharmaceuticals, a domain in which the FDA generally exerts considerable influence over industry. However, the FDA has direct oversight over specific implementations of medical AI systems in its regulation of medical devices; the Ada Lovelace Institute's report *Safe before Sale* gives an overview of the FDA's medical-device process and lessons for AI governance.⁶

II. History of the FDA as a History of Public Scandals

The pharmaceutical sector is a particularly high-investment industry, and one in which products pose high levels of risk both to individuals and to public health. Historically, the production and sale of drugs was largely unregulated until the late nineteenth century, and much of this nascent drug industry was opaque to consumers. The Department of Agriculture's Division of Chemistry was the primary predecessor to the FDA,⁷ but a series of scandals—followed by the passage of legislation by Congress—set the stage for stronger regulatory intervention, including the following:

- **1906 Pure Food and Drug Act.** This was the first of a series of laws passed in response to concerns about “patent medicine” and the use of dangerous ingredients in drugs sold to consumers, revealed through muckraking exposés by journalists. It assigned to the Division of Chemistry responsibilities for enforcing labeling requirements for drugs, whereby variations from a standard form had to be acknowledged on the label.⁸ It is also widely considered to be the law establishing the FDA, though the agency was not known by that name until almost 25 years later.
- **1938 Food, Drug, and Cosmetics Act.** This legislation was passed in response to the scandal of sulfanilamide elixir poisoning causing the deaths of more than a hundred people, including children.⁹ The Act prohibited drugs from being marketed unless the manufacturer submitted an application to the FDA proving their safety.
- **1962 Amendment to the Food, Drug, and Cosmetics Act.** This amendment introduced efficacy into the FDA's purview, following the thalidomide disaster in which a

⁵ Office of the Commissioner, “What We Do,” Food and Drug Administration, November 15, 2023, <https://www.fda.gov/about-fda/what-we-do>.

⁶ Merlin Stein and Connor Dunlop, *Safe before Sale: Learnings from the FDA's Model of Life Sciences Oversight for Foundation Models*, Ada Lovelace Institute, December 13, 2023, <https://adalovelaceinstitute.org/report/safe-before-sale>.

⁷ John P. Swan, “How Chemists Pushed for Consumer Protection: The Food and Drugs Act of 1906,” *Chemical Heritage* 24, no. 2 (Summer 2006): 6–11, <https://www.fda.gov/files/about%20fda/published/How-Chemists-Pushed-for-Consumer-Protection--The-Food-and-Drugs-Act-of-1906.pdf>.

⁸ This early authority gave the FDA the ability to regulate claims that manufacturers made about drugs. Because this model proved inadequate to protect the public, the FDA was given greater authority over time to regulate the products themselves. (Thanks to Chris Morten for this note.) See “Part I: The 1906 Food and Drugs Act and Its Enforcement,” Food and Drug Administration, April 24, 2019, <https://www.fda.gov/about-fda/changes-science-law-and-regulatory-authorities/part-i-1906-food-and-drugs-act-and-its-enforcement>.

⁹ Philip R. Lee and Jessica Herzstein, “International Drug Regulation,” *Annual Review of Public Health* 7 (1986): 217–35, <https://www.annualreviews.org/content/journals/10.1146/annurev.pu.07.050186.001245>.

drug designed to reduce morning sickness in pregnant people led to life-threatening birth defects. The FDA's refusal to approve the drug, previously approved in other countries, "deeply shaped the public's perception of the Agency, and helped justify significant expansions in its regulatory authority."¹⁰

- **2007 Food and Drug Administration Amendments Act.** This bill introduced a number of amendments, most significantly by mandating transparency both from the FDA and from regulated companies. It required that the FDA share the analysis underlying its decisions to approve new drugs, and that drug companies share some of their own data with the public.¹¹ This helped foster an information ecosystem that exists both within and beyond the FDA: companies must generate information on their products to share with the FDA, and both parties must share some of this information with the broader public.¹²
- The FDA has undergone many other administrative, funding and process adjustments, including:
 - **The 1992 Prescription Drug User Fee Act (PDUFA).** This legislation introduced "user fees" whereby industry would directly fund the FDA (see [industry capture](#) for more info).¹³
 - **The 21st Century Cures Act.** The act responds to pressure from industry and Congress to speed the approvals process by establishing accelerated approvals pathways, including the provision of the results of observational studies in some circumstances in place of full **clinical trials**, under certain circumstances (see the [premarket assessment section](#) for more info).¹⁴
- The COVID-19 pandemic highlighted the way issues regarding medical products can become politicized, and led to discussion about how to separate the agency's scientific evaluation from heated policy debates, prompting some to suggest the FDA should be structured as an **independent agency**.¹⁵

¹⁰ Amy Kapczynski, "Dangerous Times: The FDA's Role in Information Production, Past and Future," *Minnesota Law Review* 102 (July 2018), <https://scholarship.law.umn.edu/mlr/130>.

¹¹ See Christopher J. Morten, Gabriel Nicholas, and Salomé Viljoen, "Researcher Access to Social Media Data: Lessons from Clinical Trial Data Sharing," *Berkeley Technology Law Journal* 39, no. 1 (April 2024): 109–204, <https://doi.org/10.15779/Z38QF8JK81>; Matthew Herder, Christopher J. Morten, and Peter Doshi, "Integrated Drug Reviews at the US Food and Drug Administration—Legal Concerns and Knowledge Lost," 180(5) *JAMA Intern Med* 629-30 (March 2, 2020), doi:10.1001/jamainternmed.2020.0074; and Food and Drug Administration, "FY2025 FDA Budget Summary."

¹² Thanks to Chris Morten for this point.

¹³ Maryanne Demasi, "From FDA to MHRA: Are Drug Regulators for Hire?" *BMJ* 377 (June 29, 2022): o1538, <https://doi.org/10.1136/bmj.o1538>.

¹⁴ Office of the Commissioner, "21st Century Cures Act," Food and Drug Administration, January 31, 2020, <https://www.fda.gov/regulatory-information/selected-amendments-fdc-act/21st-century-cures-act>.

¹⁵ Laura Karas, "FDA's Revolving Door: Reckoning and Reform," *Stanford Law and Policy Review* 34, no. 1 (2023): 1–66, <https://law.stanford.edu/publications/fdas-revolving-door-reckoning-and-reform>. A similar set of discussions ensued following the decision by Health and Human Services Secretary Kathleen Sebelius to override the FDA's decision to make emergency contraception more widely available. See Gardiner Harris, "Plan to Widen Availability of Morning-After Pill Is Rejected," *New York Times*, December 7, 2011, <https://www.nytimes.com/2011/12/08/health/policy/sebelius-overrules-fda-on-freer-sale-of-emergency-contraceptives.html>.

Section 2: What Does the FDA Do?

Public debates about an “FDA for AI” often operate in a broad analogical manner, asking what a federal agency regulating AI would look like. Yet this is a blunt instrument for a conversation deserving of greater nuance.

As a productive starting point for our analysis, we chose to identify the key regulatory functions for the FDA and assess their relevance for AI. Using thematic coding, we bucketed these regulatory functions into three groupings, analyzing how each functions within the context of the FDA, and how each has considerations for artificial intelligence—points of alignment, departure, or other factors that surfaced from engaging with the FDA example. We’ve summarized these takeaways below, and offer a more thorough accounting in Appendix 2.

I. Premarket Approval

What is it?

“**Premarket approval**” refers to the FDA’s role in scrutinizing and approving (or refusing to approve) prescription drugs *before* they enter widespread circulation. This process gives the FDA control over how a drug can be marketed—both to the public and to the physicians who prescribe drugs to the public.¹⁶ Much of the FDA’s power therefore lies in its gatekeeping function. It controls a key gateway through which pharmaceuticals need to pass in order to enter the market, providing strong incentives for companies to comply with the approvals process.

As the central regulatory authority responsible for evaluating drugs before they are marketed, the FDA examines both the **safety** of drugs and their **efficacy**, making an approval determination based on a risk-benefit analysis.

These two qualities—safety and efficacy—do not exist in full isolation from each other, and at times may in fact be in tension. A more effective drug may introduce increased risks of side effects, and regulators must evaluate whether the benefits of tackling a specific disease outweigh the ancillary harms that are introduced. Only by considering safety and efficacy together is it possible for the agency to make meaningful assessments of whether the drug should be allowed on the market.

In the pharmaceutical context, the FDA uses a benchmarking process that is both flexible and standardized in the form of **end points**: evaluative metrics tailored to an indication of disease that are not only valid and generalizable, but also reflect a particular outcome being measured. End points are established in agreement between FDA staff and drug manufacturers, and form

¹⁶ Center for Drug Evaluation and Research, “Prescription Drug Advertising | Questions and Answers,” Food and Drug Administration, updated June 19, 2015; accessed July 10, 2024, <https://www.fda.gov/drugs/prescription-drug-advertising/prescription-drug-advertising-questions-and-answers>.

a key part of the clinical trial process as determinants of whether the drug has successfully achieved a stated health outcome.¹⁷ **Surrogate end points** serve as proxies, metrics that are closely linked to more traditional end points but that may enable swifter evaluation by substituting a short-term outcome for a long-term one. (For example, one workshop participant referenced reduction in tumor size as an example of a surrogate end point that is clinically verifiable on a shorter time frame than seeing a patient's cancer go into remission).¹⁸

Regardless of the pathway used, the premarket approach to drug approvals necessitates that drug manufacturers document their reasoning and decision-making throughout the development process. This feeds into another important function of the FDA—that of [producing information and expertise](#)—which we discuss at greater length below.

How could it work for AI?

At present, with few exceptions, many artificial intelligence systems do not go through any kind of standardized evaluation process prior to entering commercial deployment in the United States (though some AI systems do go through post-market evaluation; see Appendix 1). Under a **“permissionless innovation”** approach, regulators have taken a light-touch approach to the market, tending to step in **ex post** to correct harms after they have occurred.¹⁹

In practice, this *ex post* approach to regulatory enforcement is often triggered by negative press coverage, independent auditing by researchers or journalists, whistleblowing, or consumer reporting. These triggers can lead enforcement agencies to open up an investigation, evaluate compliance with the contours of existing law, and, if merited, institute penalties for failure to comply with legal mandates.

This approach has a number of weaknesses: it tasks under-resourced actors with the most onerous elements of accountability. It means that redress is often unevenly distributed, since garnering sufficient attention to merit redress often requires access to public platforms. It means that evaluations of legal compliance are often ad hoc and highly context-specific, in the absence of more objective measures carried out across the board. And often, the enforcement process occurs long after the harm has been incurred, too late to effectively remediate.

The FDA's premarket review for drugs offers useful insights for the AI policy community to consider:

1. Premarket review places the burden on the manufacturer to prove a product is safe, rather than on the public or enforcement agencies to identify instances where harm has occurred.

¹⁷ Charlie McLeod et al., “Choosing Primary Endpoints for Clinical Trials of Health Care Interventions,” *Contemporary Clinical Trials Communications* 16 (December 2019): 100486, <https://doi.org/10.1016/j.conctc.2019.100486>.

¹⁸ Ibid.

¹⁹ Darrell M. West, “The End of Permissionless Innovation,” Brookings, October 7, 2020, <https://www.brookings.edu/articles/the-end-of-permissionless-innovation>.

2. The level of flexibility adopted in the FDA model could offer a mechanism through which the rigor of the evaluation process is calibrated to the specific domain of deployment, mitigating assertions that FDA-style regulation is overly onerous and expensive. For example, the **end-point** approach not only offers flexibility around what counts as proof of “safety and efficacy” but also shapes what type of information the FDA can ask for. This would also enable the agency to appropriately calibrate the standard-setting process to account for contextual factors important to how AI is developed and deployed in the world, and the dynamic nature of system development.
3. Rather than engaging fixed standards for evaluation of products already on the market, the FDA approach incentivizes companies to invest in the development of evaluation measures well attuned to their systems in order to “show their work.”

In contrast to the **“red-teaming”** approach that AI companies have tended to favor, this premarket approach mandates that firms do more than say, “Trust us, we’ve done our homework.” Red teaming is generally a black-boxed process. It looks solely at areas of risk or vulnerability rather than efficacy, and leaves it to companies to define the metrics and processes through which their products are tested. Moreover, red teaming does not substantiate safety, functionality, or efficacy claims, which are all required when undergoing any regulatory process. By contrast, the approach to clinical testing utilized by the FDA encourages greater transparency and documentation of development practices, raises the bar for standard setting and benchmarking while maintaining scope for flexibility attuned to context, and enables empirical evaluation not only of harm but of whether a system works in the first place.

An important consideration is the complexity of AI **supply chains**, which often string together multiple service providers or firms with multiple use cases. The FDA has a process for quality control across the entirety of the supply chain, including engaging with some regulation at the level of the suppliers of components of drugs;²⁰ its enforcement, however, is most intense at the level of a particular end application, where validating the safety and effectiveness of a given product is more manageable.

The FDA regulatory model may not be as well suited to points in the supply chain that are more distant from the context of deployment, such as the base or “foundation model” layer. It is difficult to evaluate a general-purpose system for safety and efficacy because the full range of use cases is unknown. Here, other regulatory design approaches—such as financial regulation and its treatment of systemic risk,²¹ or emissions monitoring regulation—may offer more useful corollaries.

²⁰ This regulatory process takes place via drug master files (DMFs), which may be instructive for regulating foundation models. See Food and Drug Administration, “Drug Master Files (DMFs),” November 3, 2023, <https://www.fda.gov/drugs/forms-submission-requirements/drug-master-files-dmfs>.

²¹ See Julia Black and Andrew Murray, “Regulating AI and Machine Learning: Setting the Regulatory Agenda,” *European Journal of Law and Technology* 10, no. 3 (2019), <https://www.ejlt.org/index.php/ejlt/article/view/722>; and Carsten Jung and Bhargav Srinivasa Desikan, “Artificial Intelligence for Public Value Creation: Introducing Three Policy Pillars for the UK AI Summit,” IPPR, October 25, 2023, <https://www.ippr.org/articles/ai-for-public-value-creation>.

There are nevertheless safety processes, testing, and documentation that can be mandated for all points along the supply chain. At a minimum, mandates for clear documentation of base models, including the data used to train them, will be necessary to enable evaluation at the application layer. (See the [Producing Information and Expertise](#) section for more on this.)

II. Post-Market Monitoring and Enforcement

What is it?

The FDA recognizes that simply mandating testing and documentation prior to a drug or device entering the market does not ensure safety, and that ongoing **post-market monitoring** is required to ensure that new risks are caught as they emerge.

The FDA uses a variety of tools for post-market monitoring. It benefits from what is sometimes called **passive surveillance**: the voluntary reporting of safety incidents by doctors, insurers, and pharmaceutical companies themselves. It also engages in **active surveillance**, which involves reviewing and monitoring electronic health records' data for reports of adverse reactions or harm that could be linked to the use of a particular drug.²² Because premarket approval and disclosures are paired with liability, other regulators and patients (including those in US states) can bring tort court cases, which can lead to additional discovery when a product appears to have caused patients harm.²³

The FDA's powers are weaker here in comparison to premarket approval, the most powerful stage of regulatory intervention: the period before drugs enter the market is when alignment between regulatory power and companies' incentives to comply reach their peak. Past the point of market entry, the FDA retains some ability to act in the public interest, through the measures outlined above—but we see a significant drop in the agency's ability to act and in its track record of doing so successfully.

How could it work for AI?

In both the context of the FDA and in AI, assuring downstream compliance after a product enters the market is a regulatory challenge.

Post-market surveillance for AI is particularly obstacle ridden. But this remains the dominant way we enforce laws on artificial intelligence today.

A better approach would be to more actively monitor AI systems, which are highly dynamic both in the frequency of updates to AI systems and active data flows. But how to do so effectively

²² Mary Wiktorowicz, Joel Lexchin, and Kathy Moscou, "Pharmacovigilance in Europe and North America: Divergent Approaches," *Social Science & Medicine* 75, no. 1 (July 2012): 165–170, <https://doi.org/10.1016/j.socscimed.2011.11.046>.

²³ National Academy of Engineering, *Product Liability and Innovation: Managing Risk in an Uncertain Environment* (Washington, DC: National Academies Press, 1994), <https://doi.org/10.17226/4768>.

needs more deliberation: for example, in the pharmaceutical context, regulators use electronic health records to proactively surface patterns indicating harmful effects that could be tied to a particular drug's use.²⁴ In financial regulation, mechanisms like scenario planning are frequently used to anticipate known crisis patterns before they occur. Are there equivalents that could be better leveraged to more proactively surface AI-enabled harms? (See the section on [Producing Information and Expertise](#) for more information.)

Such efforts could complement ongoing auditing and impact assessments of AI systems by independent entities; how to effectively track and mitigate harms across the development life cycle of an AI system is thus likewise important, requiring the appropriate calibration of obligations to each phase of development.²⁵ When AI systems are deployed out in the world, they are exposed to rapid changes that can alter data and processes in real time. Compared to a drug, which is a fixed object, many AI systems are inherently variable, requiring ongoing monitoring and risk mitigation. Approaches like financial regulation may be useful analogies that offer both the vocabulary and the mechanisms for evaluating risk more systematically.²⁶

Wrangling the potentially varied provenance of AI system components, and establishing a calculus that leads to an informed understanding of the point in a system's development at which the harm may have been introduced, remains a difficult task. Clearer documentation mandates could help address the muddled provenance of AI system components, but this is an issue that needs deeper deliberation. The “cascading” approach outlined by the UK's National Cyber Security Centre, in which obligations are attuned to each phase of development (including the disclosure of risks each actor was *unable* to evaluate), may also offer a useful and similarly concrete complement to FDA-style approaches.²⁷

Finally, a pressing policy question is how to address risk posed by an AI system once identified. The FDA's power is strengthened by the heft of the penalties it can leverage: its ability to levy

²⁴ This access is provided through surveillance capacities enabled by the owners of these systems, such as health systems. For more, see the Food and Drug Administration (FDA) Sentinel Initiative, <https://www.sentinelinitiative.org>.

²⁵ See Ian Brown, “Expert Explainer: Allocating Accountability in AI Supply Chains,” Ada Lovelace Institute, June 29, 2023, <https://www.adalovelaceinstitute.org/resource/ai-supply-chains>; Harry Farmer, “Regulate to Innovate,” Ada Lovelace Institute, November 29, 2021, <https://www.adalovelaceinstitute.org/report/regulate-innovate>; and Pegah Maham and Sabrina Küspert, “Governing General Purpose AI — A Comprehensive Map of Unreliability, Misuse and Systemic Risks,” *interface*, July 20, 2023, <https://www.stiftung-nv.de/de/publikation/governing-general-purpose-ai-comprehensive-map-unreliability-misuse-and-systemic-risks>.

²⁶ See, e.g., Financial Stability Oversight Board, “Analytic Framework for Financial Stability Risk Identification, Assessment and Response,” Federal Register 88, no. 218, November 14, 2023, <https://home.treasury.gov/system/files/261/Analytic-Framework-for-Financial%20Stability-Risk-Identification-Assessment-and-Response.pdf>; U.S. Department of the Treasury, “The Financial Services Sector's Adoption of Cloud Services,” n.d., accessed July 10, 2024, <https://home.treasury.gov/system/files/136/Treasury-Cloud-Report.pdf>; and Danny Brando et al., “Implications of Cyber Risk for Financial Stability,” Federal Reserve, May 12, 2022, <https://www.federalreserve.gov/econres/notes/feds-notes/implications-of-cyber-risk-for-financial-stability-20220512.html>.

²⁷ National Cyber Security Centre, “Guidelines for Secure AI System Development,” November 27, 2023, <https://www.ncsc.gov.uk/collection/guidelines-secure-ai-system-development>.

huge fines, strip certification, and refuse to approve future products is a powerful deterrent for negligent behavior, particularly since the companies making drugs generally need to bring new products to the agency for approval in the future. There are formal mechanisms for downstream accountability, such as recalling products after the fact, though the FDA's ability to enact these remedies is weakened once they are in commercial use. Companies also remain liable for harms caused to the public after drugs are made available for wide release.

Currently, the bulk of regulatory enforcement of existing law in AI occurs *ex post*, and is thus subject to these challenges; even identifying where AI systems are currently in use remains a significant gap. In addition, establishing liability and then demonstrating causation in the AI context are significant barriers.²⁸

III. Producing Information and Expertise

What is it?

Pharmaceuticals offer an interesting analogy to AI because, with an eye to history, drugs were similarly cryptic and underscrutinized at the time the FDA was formed. Amy Kapczynski has written extensively about how the FDA played an important role in motivating the production of information that has reduced the opacity of pharmaceuticals, contributing to our knowledge of how drugs work—with benefits that accrue to the entire sector.²⁹ The FDA thus acts as a key conduit for information to be conveyed to the public, as well as to relevant expert stakeholders who seek to represent the interests of the public (the research community and auditors, patient advocates, other regulators, clinicians).

How could it work for AI?

Many elements of artificial intelligence are currently opaque and underscrutinized, either due to corporate secrecy or because of endemic challenges to interpreting and explaining the outputs of AI systems.³⁰ This presents a significant challenge in regulating artificial intelligence: the absence of key information about how AI systems are constructed and how they function hinders the effectiveness of auditing, benchmarking, and validation.³¹ Some of this opacity can be attributed to misaligned incentive structures; left to their own devices, companies are simply not well placed or well motivated enough to consider or prioritize certain questions.³² And other

²⁸ See Mihailis Diamantis, “Vicarious Liability for AI” *Indiana Law Journal* 99, no. 1 (Winter 2023): 317–334, . <https://papers.ssrn.com/abstract=3850418>; and Miriam Buiten, Alexandre de Streel, and Martin Peitz, EU Liability Rules for the Age of Artificial Intelligence, Centre on Regulation in Europe, March 2021, <https://doi.org/10.2139/ssrn.3817520>.

²⁹ Amy Kapczynski, “Dangerous Times.”

³⁰ Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Cambridge, MA: Harvard University Press, 2016), <https://www.hup.harvard.edu/books/9780674970847>.

³¹ *Ibid.*

³² Christopher Morten, “Publicizing Corporate Secrets,” *University of Pennsylvania Law Review* 171, no. 5 (January 1, 2023): 1319. <https://doi.org/10.58112/plr.171-5.2>.

questions may indeed prove intractable empirical barriers to our ability to understand AI systems.³³

A number of these information-generating approaches are directly applicable to AI systems. The power to elicit the necessary information to effectively evaluate an AI system, to require AI systems to be labeled as such, and to report adverse events are all within the scope of existing AI governance proposals.³⁴ These seek to strengthen the baseline offered under existing law: in its Executive Order on AI, the White House included the use of the Defense Production Act to elicit information from AI companies about how they are evaluating the safety of their systems (above a scale threshold) and report this information to the Department of Commerce.³⁵ Moreover, the Federal Trade Commission (FTC)’s existing authorities on substantiation in advertising enable the agency to request information from AI companies asking them to validate claims they make publicly about the capabilities of their product offerings.³⁶

A key distinction is that the “user” of an AI system—the entity procuring AI—is often not the same as the entity on which the system is used.

This distinction matters tremendously. Disclosures and other transparency mechanisms may be important for the decision-making of entities about whether and under what conditions to use AI, but they often will be insufficient to enable those on whom AI is used to remediate harm. Often, individuals are not in a position to make meaningful decisions about how AI impacts their lives. In contexts like hiring, insurance, healthcare, education, and finance, members of the public rarely are given the opportunity to shape how AI may be used in a decision-making process even as it significantly impacts their access to resources and life chances—and they aren’t given the autonomy to opt out. Paying attention to the effects of information and power asymmetries will be particularly important in AI governance; those involved in regulatory design should remain vigilant about how to implement mechanisms that would function more meaningfully in the public’s interest.

Section 3: Lessons from the FDA Model

What lessons can we learn from the FDA model for AI governance, and how might FDA-style interventions be applied to AI? This section outlines three crosscutting themes from the experience of the FDA’s regulation of pharmaceuticals. Rather than homing in on specific

³³ Zachary C. Lipton, “The Mythos of Model Interpretability,” arXiv, March 6, 2017, <https://doi.org/10.48550/arXiv.1606.03490>.

³⁴ See examples in [Appendix 1](#).

³⁵ White House, “Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence,” October 30, 2023, <https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence>.

³⁶ Federal Trade Commission, “FTC Policy Statement Regarding Advertising Substantiation,” June 24, 2014, <https://www.ftc.gov/legal-library/browse/ftc-policy-statement-regarding-advertising-substantiation>.

functions like the previous section did, this one looks at how these functions interact to produce particular outcomes in the sector.

I. Establishing Efficacy and Safety

The safety and efficacy of products must be evaluated in parallel to make a fulsome assessment of their impact on society at large. In the context of AI, policymaking and regulatory activity have tended to index heavily on risk, and have insufficiently evaluated the efficacy of AI systems.

Establishing efficacy in the pharmaceutical context is a complex task: the history of the FDA is a history of many value-laden disputes around how the effects of drugs ought to be measured and what ought to count as evidence. Compared to AI, though, the task of pharmaceutical regulators is comparatively straightforward: Does the drug work when evaluated against an agreed-upon end point?³⁷

While evaluating the claims made by companies themselves offers a useful starting point, doing so can be much more complicated in the context of artificial intelligence, where systems are highly complex and also sociotechnical—a system that operates effectively in controlled environments may fail to function appropriately when deployed in the real world.³⁸ The makers of general-purpose AI systems are also much less likely to make specific claims against which efficacy could readily be tested.

Furthermore, some AI systems are *deterministic* (i.e., we can with a reasonable degree of accuracy understand and trace how an outcome was defined and arrived upon), while others are *probabilistic* (we can understand the basic mechanisms of how the system functions, but it may be difficult or impossible to consistently explain retrospectively, or anticipate proactively, how it will behave). In addition, the measures needed to evaluate effectiveness in AI are inherently more dynamic, multivariate, and complex than in the pharmaceutical context: they require contextual expertise from across a range of sectors, not solely constrained to technical expertise but incorporating the domains in which an AI system is deployed.³⁹

³⁷ Within this, there are value-laden disputes throughout the FDA’s history around measurements. (What should count as evidence?)

³⁸ Inioluwa Deborah Raji, I. Elizabeth Kumar, Aaron Horowitz, and Andrew D. Selbst, “The Fallacy of AI Functionality,” *FaccT ’22: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (June 2022): 959–72, <https://doi.org/10.1145/3531146.3533158>.

³⁹ Michael Feffer, Michael Skirpan, Zachary Lipton, and Hoda Heidari, “From Preference Elicitation to Participatory ML: A Critical Survey & Guidelines for Future Research,” *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society* (August 2023): 38–48, <https://doi.org/10.1145/3600211.3604661>; Fernando Delgado, Stephen Yang, Michael Madaio, and Qian Yang, “The Participatory Turn in AI Design: Theoretical Foundations and the Current State of Practice,” arXiv, October 2, 2023, <https://doi.org/10.48550/arXiv.2310.00907>; Abeba Birhane et al., “Power to the People? Opportunities and Challenges for Participatory AI,” *EAAMO ’22: Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization* (October 2022): 1–8, <https://doi.org/10.1145/3551624.3555290>; Laura Weidinger et al., “Sociotechnical Safety Evaluation of Generative AI Systems,” arXiv, October 31, 2023, <https://doi.org/10.48550/arXiv.2310.11986>.

In some settings, we may be willing to take on this degree of uncertainty—it's sufficient to know that an AI system works well enough. In others, it may be necessary to have a high degree of certainty both that the system works effectively and that it will not behave in ways that are detrimental to particular users or to society at large. Without an approach to benchmarking and validation of AI that considers safety and efficacy in tandem, we lack the necessary information to make these kinds of decisions.

There will always be potential harms from AI; the regulatory question thus must consider whether the benefits outweigh the harms. But to know that, we need clear evidence—which we currently lack—of the specific benefits offered by AI technologies.

To serve the public interest, measures of efficacy should be considered carefully. They should not be primarily or solely indexed on profit or growth, but should take into account benefits to society more generally.

Regulatory approaches in AI should require developers of AI systems to explain how an AI system works, the problems it attempts to address, and the benefits it offers—not just evaluate where it fails. Accurately measuring and validating noneconomic benefits has become a key challenge in other domains (notably in the context of carbon emissions reduction targets), and developing robust metrics for this should be a priority for AI governance.

II. From Opacity to Openness

The FDA model offers a powerful lesson in transparency: product safety cannot be divorced from the process of optimizing regulatory design for information production. Prior to the existence of the agency, much of the pharmaceutical industry was largely opaque, in ways that bear similarities to the AI market.

Over time, the FDA's interventions have expanded the public's understanding of how drugs work by ensuring firms invest in research and documentation. Beyond simply understanding incidents in isolation, it has catalyzed and organized an entire field of expertise and disseminated this expertise across stakeholders, enriching our understanding of pharmaceuticals and their role in our society and economy.⁴⁰

This information-production function is particularly important for AI. Key players in the market are incentivized against transparency, and even identifying these actors in the first place is a challenging task absent regulatory intervention.

Many specific aspects of information exchange in the FDA model offer lessons for thinking about AI regulation. For example, in the context of pharmaceuticals, there is a focus on multistakeholder communication that requires ongoing information exchange between staff, expert panels, patients, and drug developers. Drug developers are mandated to submit troves of internal documentation, which the FDA reformats for public release. The FDA also manages a

⁴⁰ Kapczynski, "Dangerous Times."

database of adverse incidents, clinical trials, and guidance documentation.⁴¹ What's more, it produces its own independent analysis on top of industry-supplied data that offers an important check that sometimes differs from—and even challenges—industry conclusions.⁴²

Implementing documentation requirements of this sort for AI would represent a significant change from the current accountability vacuum in AI. Encouraging AI firms to adopt stronger monitoring and compliance activities like recordkeeping and documentation practices would substantively change those firms' approach to building systems and potentially even their operating models. Such monitoring and compliance may also need to extend to the agency itself, ensuring that agency leadership doesn't use opacity to limit second-guessing of its decision-making, establishing information law rules that govern information flows between the regulator, researchers, and the broader public. Moreover, traceability remains an underexplored field in the context of AI: change and control management systems in software required significant investment, and similar investment in corollary approaches in AI would alleviate the burden on smaller players.⁴³

On one hand, this may make the development process more expensive and difficult, requiring additional documentation and validation processes rather than encouraging open experimentation.⁴⁴ On the other, such mandates would have beneficial effects for AI governance by streamlining organizational processes, ensuring durability of knowledge of how systems were developed and creating greater internal transparency and accountability. Requiring that companies conduct baseline measures to adequately scrutinize and document the development process would also enable and increase the effectiveness of external auditing, facilitating the development of an "ecosystem of inspection." It would also provide legal hooks for *ex post* enforcement, and aid the work of enforcement agencies when they need to investigate AI companies.

It is worth considering how such measures may differentially impact companies of various sizes and stages of development: AI startups may express more hesitancy about potential advantages that could be gained by bigger companies when they provide information about an AI product still in development. Thus it would be worth considering in more depth the right balance between publishing information that enables public validation of the outcomes of assessments and directly publishing the information provided by companies; the pharmaceutical

⁴¹ Center for Drug Evaluation and Research, "FDA Adverse Event Reporting System (FAERS) Public Dashboard," Food and Drug Administration, December 7, 2023, <https://www.fda.gov/drugs/questions-and-answers-fdas-adverse-event-reporting-system-faers/fda-adverse-event-reporting-system-faers-public-dashboard>.

⁴² See Christopher J. Morten and Amy Kapczynski, "The Big Data Regulator, Rebooted: Why and How the FDA Can and Should Disclose Confidential Data on Prescription Drugs and Vaccines," *California Law Review* 109 (April 2021), <https://www.californialawreview.org/print/the-big-data-regulator-rebooted-why-and-how-the-fda-can-and-should-disclose-confidential-data-on-prescription-drugs-and-vaccines>; and Kapczynski, "Dangerous Times."

⁴³ Thanks to Heidy Khlaaf for this point.

⁴⁴ Emily Black et al., "Less Discriminatory Algorithms," *Georgetown Law Journal* 113, no. 1 (2024), <https://doi.org/10.2139/ssrn.4590481>.

process may offer useful points of comparison but this will likely need further study in the specific context of artificial intelligence.

III. Generating Agency and Rebalancing Power

The FDA approach creates and distributes new sites of agency in the healthcare system, empowering actors like doctors who are obligated to work in their patients' best interest.

The power of FDA regulation comes in part from other actors in the system, including physicians, insurers, whistleblowers, and other actors who strengthen its monitoring regime. This has acted as an important second line of defense in pharmaceuticals, where the regulatory process has been insufficiently rigorous.

By contrast, we lack similar professional obligations in the AI context, dependencies and sites of friction remain comparatively immature, and the relevant actors are not necessarily incentivized toward accountability. AI is frequently used in ways that are designed to ultimately undermine the agency of those on whom it's used.

In comparison to pharmaceuticals, where numerous interdependencies exist across actors within a market—including doctors, insurance companies, pharmaceutical companies, and patient advocacy groups—artificial intelligence has relatively few. The dependencies that do exist tend to be unidirectional: smaller companies and startups depend on the resources dominated by cloud infrastructure providers like Google, Amazon, and Microsoft.

An important distinction brought up throughout this report is that the “users” of AI systems are often not the group most affected by those systems. In many instances, AI is used by comparatively powerful entities—employers, law enforcement agencies, healthcare providers, banks—on those who are less powerful.⁴⁵ Often, those on whom AI is used are not informed of its use; even if they are informed, they may lack the necessary information, or the power and authority, to seek redress if the system has caused harm.

These points of distinction between AI and pharmaceuticals, or many other consumer products for that matter, merit further attention. In particular, such differences raise questions about how information should be redistributed not only among the many actors involved in the development, procurement, and regulation of AI systems, but also among those on whom AI systems are used. It also raises questions about what sorts of responsibilities and obligations different actors should have to mitigate the power imbalances that characterize much of AI use and deployment, and how to effectively introduce the necessary friction to ensure these responsibilities are carried out.

⁴⁵ See “Algorithmic Management: Restraining Workplace Surveillance,” AI Now Institute, April 11, 2023, <https://ainowinstitute.org/publication/algorithmic-management>; AI Nationalism(s) Executive Summary; and Philip Alston, “Report of the Special Rapporteur on Extreme Poverty and Human Rights,” October 11, 2019, https://srpoverty.org.files.wordpress.com/2019/10/a_74_48037_advanceuneditedversion-1.pdf.

Section 4: Challenges for FDA-Style Interventions

I. Establishing a Regulatory Perimeter

FDA regulation for pharmaceuticals is triggered by the “marketing” of a drug as a critical gate to entry. In other industries, there are gates around the sale of certain products, which may be preferable over marketing given First Amendment concerns (see Section III below). Any attempt at sector-specific AI regulation will run into a thorny set of definitional questions: What constitutes the AI market, and how do products enter into commercial use?

In the absence of a regulatory mandate that requires AI companies to come forward and declare themselves, the contours of the “AI market” are vaguely defined: Is every company that develops an AI system internally part of the market and thus open to scrutiny? If so, then companies like Walmart are AI companies. By contrast, many companies do not develop AI systems themselves, but procure and deploy systems developed by others, potentially fine-tuning models or making other adjustments in the process.

A second question related to perimeter concerns scale and impact: an approach taken in the Executive Order on AI and in the White House Voluntary Commitments relies on a scale threshold—systems trained on computational power measured at 10^{26} floating-point operations per second (FLOPS)—to carve out systems for scrutiny, based on the presumption that particular harms are associated with especially large systems.⁴⁶ However, this presumption deserves closer scrutiny: the effect of this approach is to exclude all systems currently in operation from such reporting requirements, as this threshold just exceeds the largest currently active model. Furthermore, this approach overlooks the fact that the risks associated with AI depend closely on the contexts within which those systems are deployed.⁴⁷

This is not the only method for scoping AI systems for scrutiny. Other approaches that surfaced throughout the expert convening include examining systems based on a set of risk classifications, similar to the approach taken under the EU’s AI Act (see Appendix 1 for a list of other proposals incorporating risk classifications); structuring the market based on commercialization of an AI product; or adopting a supply-chain approach that identifies different sets of actors involved in different phases of AI development and adopting governance mechanisms tailored to the specifics of their roles.⁴⁸

⁴⁶ This scale hypothesis is articulated in Markus Anderljung et al., “Frontier AI Regulation: Managing Emerging Risks to Public Safety,” arXiv, November 7, 2023, <https://doi.org/10.48550/arXiv.2307.03718>; and previously by Lennart Heim in “The Case for Pre-Emptive Authorizations for AI Training,” June 10, 2023, <https://blog.heim.xyz/the-case-for-pre-emptive-authorizations>.

⁴⁷ Heidy Khlaaf, “Toward Comprehensive Risk Assessments and Assurance of AI-Based Systems,” Trail of Bits, March 7, 2023, https://www.trailofbits.com/documents/Toward_comprehensive_risk_assessments.pdf.

⁴⁸ See Matt Davies and Michael Birtwistle, “Seizing the ‘AI Moment’: Making a Success of the AI Safety Summit,” Ada Lovelace Institute, September 7, 2023, <https://www.adalovelaceinstitute.org/blog/ai-safety-summit>; and Elliot

A third and final consideration relates to how a new agency would interact with existing agencies. AI systems have use cases across the economy (public, private, and defense); therefore, existing agencies have the authority to study and regulate specific applications of AI systems. A brief summary of some of those authorities is outlined in Appendix 1 below. For example, the FTC has clarified the ways their existing unfair and deceptive acts authority pertains to AI systems; the FDA has recently issued guidance for AI in medical devices; and both the Trump and Biden administrations have asked agencies to explore ways to leverage AI systems to fulfill their mandates. The creation of a new agency will require the boundaries of the agency's jurisdiction to be defined. In the case of AI, this will likely mean that projects and responsibilities within existing agencies may need to be transferred, or that interagency collaboration will need to be established.

II. Enabling Robust Enforcement

Over the past two decades, artificial intelligence development has proceeded with comparatively little regulatory scrutiny, and many firms have amassed such sufficient size and scale that the penalties of enforcement agencies like the FTC amount to little more than a budget line. Determining how to create incentive structures and sufficient regulatory friction to incentivize compliance remains a difficult regulatory design problem.

Premarket approval authority. The FDA model hinges on the FDA's ability to prevent pharmaceutical companies from marketing drugs to physicians—without which they cannot sell their drugs on the market. Controlling this essential gate to market entry is what grants the FDA a big stick, critical to its effectiveness as a regulator. At present, the analogous gates to market entry in AI are haphazard (for example, adoption of cloud services that hold federal data must go through the FedRAMP certification process, but such measures do not extend to the sector at large. OMB is also considering fast tracking certain forms of AI through the FedRAMP certification process through the Emerging Technology Prioritization Framework⁴⁹).

To have teeth, any regulatory intervention targeting the AI sector must be able to meaningfully challenge some of the biggest companies in the world. In addition to the FDA's recall and debarment powers (described in Section 1 above) there are a number of powers common to federal agencies that FDA-style interventions might draw on, including but not limited to the following:

Rulemaking. Legislation is often structured around high-level mandates and principles, and technical details are left to agency rulemaking. Most agencies undertake rulemaking following the Administrative Procedure Act (APA), which mandates a series of notice and comment

Jones, "Foundation Models in the Public Sector," Ada Lovelace Institute, accessed May 10, 2024, <https://www.adalovelaceinstitute.org/project/foundation-models-gpai/>.

⁴⁹ FedRAMP, "FedRAMP's Emerging Technology Prioritization Framework," January 26, 2024, https://www.fedramp.gov/assets/resources/documents/FedRAMP_DRAFT_Emerging_Technology_Prioritization_Framework.pdf.

periods on draft rules and judiciary review.⁵⁰ Agency rules are legally binding. Rulemaking is used across agencies (most notably the FDA) to create standards for testing, reporting and audits, incident reporting and monitoring, bright-line rules regarding product use, and programs to monitor supply chains. Flexible measures, including non-binding voluntary guidance and policy statements, may also prove important mechanisms given the dynamism of the field.

Federal Advisory Councils. The Federal Advisory Committee Act enables agencies to grant authority to “create advisory committees when nonfederal input is beneficial for decision-making.”⁵¹ This may be important to the creation of committees to assist with standards, recommendations on rules and guidance, or other input from communities most at risk from use of AI systems.

Investigation authority. Federal agencies have varying levels of authority related to investigations, ranging from inquiring about illegal activity to more specific authorities aimed at general information gathering. US enforcement agencies such as the FTC and DOJ have subpoena authority and ability to bring “civil investigation demands” (CIDs). Both subpoenas and CIDs may be used to obtain existing documents or oral testimony. Agencies can also be given investigative powers to conduct studies. For example, Section 6(b) of the FTC Act enables the Commission to conduct wide-ranging studies that do not have a specific law enforcement purpose. These types of investigations give the FTC access to nonpublic information to understand an industry.

The final category of investigative authority that a new agency might require involves the capacity to conduct interagency investigations. This entails the authority to exchange confidential information with other relevant enforcement agencies within specified limitations and confidentiality assurances. This facilitates collaboration between the FTC and other law enforcement entities, reducing the risk of redundant investigations.

Debarment. The FDA has the authority to prohibit specific individuals or corporations from engaging in FDA-regulated activities (essentially ending their career) based on illegal conduct (e.g., a clinical investigator who falsifies records). Debarment can be permanent or for a set period of time. The FDA maintains a public list of debarred entities.⁵²

Import Alerts. The FDA has the authority to issue import alerts or the ability to “refuse admission” to the US market if products “appear, from sample or otherwise” (“otherwise” may include a history of violations or a failed facility inspection) to violate the Federal Food, Drug,

⁵⁰ Congressional Research Service, “Judicial Review under the Administrative Procedure Act (APA), December 8, 2020, <https://crsreports.congress.gov/product/pdf/LSB/LSB10558>.

⁵¹ Congressional Research Service, “Federal Advisory Committee Act (FACA): Committee Establishment and Termination,” October 19, 2023 <https://crsreports.congress.gov/product/pdf/IF/IF12102>.

⁵² Food and Drug Administration, “FDA Debarment List (Drug Product Applications),” updated June 13, 2024, <https://www.fda.gov/inspections-compliance-enforcement-and-criminal-investigations/compliance-actions-and-activities/fda-debarment-list-drug-product-applications>.

and Cosmetic Act (FD&C Act).⁵³ The product is put on an import alert list to notify border officials that products should be automatically detained. Once products are detained, the owner can testify that the products are safe / abide by FDA regulations and the FDA can follow up with a determination to permit or refuse entry.⁵⁴

III. Navigating the Courts

A number of legal challenges are particularly likely to come up in any conversation about the establishment of a regulatory agency devoted to artificial intelligence. While we kept those bracketed from the scope of our main conversation, we outline some of them here:

1. First Amendment and Content Moderation

Content moderation is the term used to describe the decisions, processes, and practices that online platforms put in place regarding the treatment of “user-generated content” they host or amplify. While content moderation is often associated with social media, the overlap with AI should not be ignored. Developers of AI systems including large language models (LLMs) may be considered to be engaged in activity adjacent to content moderation, for example by making decisions regarding what data to exclude from training sets and what user prompts to block or provide fixed answers to. Developers may also choose to directly moderate the output of generative models to comport with their own platform and usage policies, attempt to address certain safety and other concerns, or minimize brand risk.

As regulators discuss analyzing/overseeing these decisions, many of the challenges that come with making objective claims of what is “safe” or “accurate” content similarly plague social media regulation. Additionally, social media platforms use AI systems to curate content in a user’s feed (e.g., Facebook’s Feed, TikTok’s For You, X’s For You) and flag or help detect content that is banned in a platform’s terms of service. Because these activities are generally understood to be editorial in nature, lawmakers’ attempts to hold platforms accountable for activities that resemble content moderation could face First Amendment challenges.

NetChoice is challenging a Texas law that prohibits social media platforms from removing or labeling user posts and requires social media companies to disclose information about how they moderate and curate user content on constitutional grounds.⁵⁵ If NetChoice wins outright, the case could create a precedent that defines government-mandated disclosures regarding online platform decisions as unconstitutional, limiting the AI policy community’s ability to mandate disclosures. Groups such as Knight First Amendment Institute are arguing that while portions of the Texas law violate the First Amendment, the law’s provisions requiring platform disclosures

⁵³ Office of the Commissioner, “Federal Food, Drug, and Cosmetic Act (FD&C Act),” Section 801(a), Food and Drug Administration, November 3, 2018,

<https://www.fda.gov/regulatory-information/laws-enforced-fda/federal-food-drug-and-cosmetic-act-fdc-act>.

⁵⁴ Congressional Research Service, “Enforcement of the Food, Drug, and Cosmetic Act: Select Legal Issues,” updated February 9, 2018, <https://crsreports.congress.gov/product/pdf/R/R43609>.

⁵⁵ *NetChoice v. Paxton*, Knight First Amendment Institute at Columbia University, accessed November 26, 2023, <http://knightcolumbia.tierradev.com/cases/netchoice-llc-v-paxton>.

should be evaluated under the legal framework set out in the Supreme Court's *Zauderer* decision, which applies deferential scrutiny to laws compelling companies to disclose factual and uncontroversial information about their services.⁵⁶ If this line of argument prevails, mandated transparency reports (and impact assessments) for platforms could remain constitutional.

2. Section 230

Section 230 of the Communications Decency Act provides a liability shield for interactive computer services (ICS) defined as “any information service, system, or access software provider that provides or enables computer access by multiple users to a computer server, including specifically a service or system that provides access to the Internet and such systems operated or services offered by libraries or educational institutions.”⁵⁷ The law specifies that no provider or user of ICS shall be treated as the publisher or speaker of any information provided by another information content provider and should not be liable for the decision to take voluntary action to restrict objectionable material.

Many AI systems may be considered interactive computer services. To the extent that an AI system is engaged in the dissemination of user-generated content, these systems could be shielded from liability for harm. Thus far, US law seems to point toward recognizing AI systems used within social media platforms as protected by Section 230.⁵⁸ Generative AI systems such as ChatGPT do not intermediate third-party content but create content, and some legal scholars believe that tools like ChatGPT will not be protected by Section 230.⁵⁹

3. US Courts and Rulemaking

As described above in Section II on enabling robust enforcement, rulemaking under the Administrative Procedure Act (APA) and other substantive laws is considered crucial for the sort of smart, effective, iterative regulation necessary for a quickly evolving administrative agency. This requires an administrative agency to be empowered to make decisions in a way that limits challenges or reversals to the extent possible. This may be rendered difficult, given the current trend of the Supreme Court seeking to limit the powers of administrative agencies.

Recently, the Supreme Court overturned the long standing principle of “Chevron deference,” derived from the famous *Chevron* case,⁶⁰ in 1984, which required courts to defer to reasonable

⁵⁶ Ibid.

⁵⁷ Congressional Research Service, “Section 230: An Overview,” updated January 4, 2024, <https://crsreports.congress.gov/product/pdf/R/R46751>.

⁵⁸ *Gonzalez v. Google*, 598 U.S. 617, May 18, 2023, https://www.supremecourt.gov/opinions/22pdf/21-1333_6j7a.pdf.

⁵⁹ Hasala Ariyaratne, “ChatGPT and Intermediary Liability: Why Section 230 Does Not and Should Not Protect Generative Algorithms,” SSRN, May 16, 2023, <https://ssrn.com/abstract=4422583>.

⁶⁰ *Chevron U.S.A., Inc. v. NRDC*, 467 U.S. 837, June 25, 1984, https://tile.loc.gov/storage-services/service/ll/usrep/usrep46_7/usrep467837/usrep467837.pdf.

agency interpretation of ambiguous statutory provisions. For several decades, this has formed the cornerstone of agency authority. *Chevron* deference made it unlikely that a pharmaceutical company would challenge, for instance, the FDA's regulation of new products and devices, clinical trials, premarket approvals etc. as discussed in Section 2, since a court of law would be likely to defer to FDA's interpretation of the relevant provisions of law. Now, pursuant to the 2024 cases *Loper Bright Enterprises v. Raimondo* and *Relentless Inc. et al. v. Department of Commerce*⁶¹ ("Loper Bright"), a reviewing court "need not and under the APA may not defer to an agency interpretation of the law simply because a statute is ambiguous."⁶² This means that when an agency interprets a statutory provision with inherent ambiguity (such as, for example, whether the FDA can regulate laboratory-developed tests as "devices"⁶³), the agency's decision could be challenged before a court of law, which will then undertake a de-novo assessment of whether the agency correctly interpreted the law. Chief Justice Roberts categorically empowered courts to take up this role, stating that "agencies have no special competence in resolving statutory ambiguities. Courts do."⁶⁴

Loper Bright continues a trend of undermining agency authority that follows the establishment of the "major questions doctrine" in *West Virginia v. EPA*⁶⁵ in 2022. There, the court stated that when a statute raises a matter of vast "economic and political significance" (i.e., a major question), then it needs to be resolved by the judiciary rather than the administrative agency, unless the agency exercise of power is supported by clear Congressional delegation.

Loper Bright does not allow the courts to overstep agency decisions in every instance. Chief Justice Roberts observes that some statutes "expressly delegate" to an agency the authority to interpret a particular statutory term. Others empower an agency to prescribe rules to "fill up the details" of a statutory scheme. The court would have to defer to such delegation by the legislature.

What does this mean for AI governance? Of course, reducing ambiguities in law is the necessary first step, since the issue of challenging agency authority arises when there is an identifiable ambiguity that the agency has stepped in to address. However, a certain degree of ambiguity is inherent in the regulation of emerging technologies, since a statute governing such technologies will necessarily need to be flexible enough to keep pace with rapidly changing technology. Thus, while setting up a new agency, legislatures should anticipate questions to be raised under *Loper Bright* as well as under *West Virginia v. EPA*, and offer the clear Congressional delegation of authority that the Supreme Court has wanted to see in both cases. To minimize litigation risk and solidify the authority of a new agency, the parent statute should

⁶¹ 603 U.S. ____ (2024), June 28, 2024, https://www.supremecourt.gov/opinions/23pdf/22-451_7m58.pdf.

⁶² *Ibid.*, p. 35.

⁶³ "FDA Takes Action Aimed at Helping to Ensure the Safety and Effectiveness of Laboratory Developed Tests," press release, Food and Drug Administration, April 29, 2024, <https://www.fda.gov/news-events/press-announcements/fda-takes-action-aimed-helping-ensure-safety-and-effectiveness-laboratory-developed-tests>.

⁶⁴ 603 U.S. ____ (2024), p. 23.

⁶⁵ *West Virginia v. EPA*, 597 U.S. 697, June 30, 2022, https://www.supremecourt.gov/opinions/21pdf/20-1530_n758.pdf.

expressly empower it to interpret any ambiguities in the law, and resolve any major questions raised within the law.

This is not foolproof, because a court could still undertake an assessment of the agency's power under the nondelegation doctrine—that is, it could assess whether the scope of delegated authority crosses into the impermissible territory of allowing an executive body to exercise essential legislative functions. It is also noteworthy that the court has recently limited powers of other agencies such as SEC and EPA, creating an expectation that the powers of the administrative state will continue to be curtailed. It is important to track these developments and design a law that plans around the new standards set by these cases.

IV. Preventing Industry Capture

High on the list of concerns about forming any novel agency charged with enforcement of an industry is the risk that commercial interests might overwhelm the regulatory authority and independence of the agency. This is particularly pressing in the context of artificial intelligence, in which the leading firms hold considerable economic and political power and a track record of increasingly assertive lobbying.⁶⁶

A prime example is in how the FDA is funded. According to one estimate from 2022, 65 percent of the FDA's work is funded through user fees that are paid for by applicant firms.⁶⁷ This takes the shape of a negotiated fee for a five-year period. The FDA negotiates with the industry it regulates for how the fees will be used: for example, when there is a medical reviewer who will be paid through the user fees of a particular applicant, that applicant will in turn receive regular reporting on how its fees were used and whether deadlines were met. This makes the FDA responsible to the companies it is reviewing for its accounting—this significantly weakens the agency's power and risks creating leverage by industry. The FDA also provides reports to public stakeholders, but there is a significant disparity between the frequency of its meetings with industry and with the public (this latter category also includes trade associations, which are typically nonprofit).

Agencies struggle to avoid becoming “captured” or overly friendly toward industry for a few reasons:

Industry lobbying. Industry players have more resources than consumer advocates to hire government affairs staff devoted to tracking the drafting of agency rules and guidance and serving on committees. This imbalance, paired with companies' ability to fund political campaigns, leads to a situation in which companies can bend text in their favor through direct engagement with agencies, or via meetings with friendly lawmakers who can write letters of support and make calls to agencies during rulemaking processes.

⁶⁶ AI Now Institute, “Tech and Financial Capital” AI Now Institute (blog), April 11, 2023, <https://ainowinstitute.org/spotlight/tech-and-financial-cap>.

⁶⁷ Demasi, “From FDA to MHRA: Are Drug Regulators for Hire?”

Revolving doors. Regulatory agencies require expert staff, which often requires hiring people who have worked in the industry being regulated. When people switch jobs, they often maintain relationships with former colleagues. Additionally, agency jobs do not pay as well as industry jobs, which creates an incentive for staff at agencies to assuage industry representatives in hopes of being recruited for jobs in the future. For example, former FDA Commissioner Dr. Scott Gottlieb joined Pfizer's board of directors within four months of announcing his resignation. As a commissioner, Gottlieb rolled out the Biosimilars Action Plan to promote the development of follow-on versions of biologic products. Pfizer happens to be a leading maker of biosimilars.⁶⁸ When agencies attempt to close the “revolving door,” it can result in an inability to hire top talent, especially in quickly evolving technical fields.⁶⁹

Consultants. In industries that rely heavily on outside evaluations/audits, the consulting industry (e.g., McKinsey, Accenture, Deloitte, EY) often becomes involved, including in direct work for expert agencies. These firms frequently conduct projects on behalf of the regulator while simultaneously working for industry players.⁷⁰

Funding dependencies. The funding model for regulatory agencies matters tremendously for its effectiveness. If a regulator is dependent on funding from industry, this can inadvertently make the regulator beholden to industry motives.

In the AI context, an additional risk emerges due to the unique structure of the sector:

Infrastructure conflict of interest. An FDA for AI will likely require access to cloud computing infrastructure in order to run audits or provision a sandbox for external audits. Unfortunately, the major providers of this infrastructure in the US (AWS, Azure, Google) will also have a suite of products that require agency oversight.

Frequent counterarguments made in response to concerns about regulatory capture are that any industry must be engaged in the regulatory conversation given that companies are closest to the ground and know the inner workings of their products. In many industries, however, regulators take a much more overtly adversarial posture expressly to ensure that the interests of the public and the economy at large are adequately protected against corporate malfeasance. A better frame might proceed from the position that any discussion about regulatory design must attend to the eventual likelihood of industry influence and must ensure, through its structure and accountability mechanisms, that industry does not get to set the agenda where it is involved in governance processes.

⁶⁸ Karas, “FDA’s Revolving Door.”

⁶⁹ Ibid.

⁷⁰ Committee on Oversight and Reform, *The Firm and the FDA: McKinsey & Company’s Conflicts of Interest at the Heart of the Opioid Epidemic, Interim Majority Staff Report*, U.S. House of Representatives, April 13, 2022, <https://oversightdemocrats.house.gov/sites/evo-subsites/democrats-oversight.house.gov/files/2022-04-13.McKinsey%20Opioid%20Conflicts%20Majority%20Staff%20Report%20FINAL.pdf>.

Conclusion

Let's return to the question that prompted this deliberation in the first place: Do we need an FDA for AI? Many of those who participated in the conversation arrived at the answer, "No, at least not in that exact form." But valuable lessons were derived from deliberating deeply on the history and regulatory functioning of a single agency, considering AI governance not from the perspective of the current status quo but instead thinking through what AI might be if regulated differently. To this end, we landed on a set of concrete takeaways, highlighted above in the executive summary, as well as a number of points that will require further deliberation.

In this report, we've sought to "show our work" so that others might be able to learn from, and build upon, our conversation. It's worth reiterating that the group did not seek to achieve consensus or any clear set of findings, nor did we arrive at these things. Instead, we engaged with the intent to have a grounded and deliberative conversation, siloed from the impetus to arrive at quick answers or to press for what's immediately possible. By muddling through, we aim to get closer to the question that should be at the heart of any conversation about AI governance: What world do we want to live in, and what role should AI play in it?

Glossary

Active surveillance. Active surveillance, in the context of FDA practices, refers to proactive collection, analysis, review, and monitoring of health data, typically in large healthcare databases, to identify adverse effects linked to the use of a particular drug.

Clinical trial. Clinical trials are research studies conducted among human volunteers to assess the safety and efficacy of new drugs, devices, or treatments.

Downstream (in foundation model supply chain). In a supply chain involving a particular foundation model, a downstream product is one that utilizes and builds on the foundation model, as opposed to an upstream product, which feeds into the foundation model in question.

End point. An end point, in the context of a clinical trial, refers to an outcome experienced by the participants of the trial that evidences the impact of the medicine or treatment being trialed, typically pre-agreed upon by the FDA and the entity running the trial. An end point can take the form of a desired clinical outcome (e.g., a reduction in blood pressure in a trial for pressure-reduction pills) or something that offers indirect evidence of a desired clinical outcome, known as a *surrogate end point*.

Effectiveness. Effectiveness, in the context of a medicine or medical treatment, measures if the drug or treatment has the real-life effect it purported to have, or represented itself as having, under controlled conditions.

Efficacy. Efficacy of a medicine or medical treatment, typically measured through an efficacy trial, refers to a determination of whether an intervention produces the expected results under ideal or controlled settings, as opposed to effectiveness trials, which measure the same in real-world settings.

Ex ante. *Ex ante* regulation refers to measures that need to be complied with prior to carrying out a regulated activity or launching a regulated product into the market, such as a license being required prior to the launch of a particular kind of AI product or a certification of process being required prior to allowing certain food products to enter the market.

[Note: This term has not been used in the report, except for in citations.]

Ex post. *Ex post* regulation governs outcomes that occur after the activity that is the subject of the regulation commences. In the context of FDA, this refers to companies being liable for harms caused to the public after drugs are commercially released; in the context of AI, it may refer to liability for harms arising as a result of the use of the AI product in question.

Foundation model. A foundation model is a large-scale, general-purpose artificial intelligence model, typically trained on vast amounts of preexisting data using unsupervised learning, that serves as a foundation for various general-purpose tasks such as generation of text, images and videos.

Investigational new drug (IND). An investigational new drug (IND) is a novel or innovative drug or biological product that does not have an approved market application from the FDA yet. It can be transported across state lines and introduced to clinical trial participants only after obtaining special authorization from the FDA for this purpose (a process known as an IND application).

[Note: This term has not been used in the report.]

New drug application (NDA). A new drug application (NDA) is the process through which sponsors of a novel drug formally propose to the FDA that the drug should be approved for sale and marketing in the US. Data gathered during clinical trials of an IND become part of this application.

Passive surveillance. Passive surveillance refers to the centralized collection, by a health authority, of reports of adverse events following the administration of a product to the population, such as a vaccine.

Permissionless innovation. In the context of product development, permissionless innovation refers to the flexibility of any person in any part of the product development process to develop new ideas for changing and improving the product, unconstrained by rules requiring prior approval. In the regulatory context, permissionless innovation refers to the idea that introduction of new technologies and innovations in the market should be permitted by default unless expressly prohibited.

Post-market monitoring. This is a system of surveillance implemented by the FDA to track adverse events related to a drug after it is marketed, in order to identify issues that did not appear during the drug-approval process.

Premarket assessment. Premarket assessment refers to the FDA's scrutiny of drugs and medical devices before they enter the market for circulation. Usually, products that are subject to such an assessment require FDA approval prior to being commercially marketed.

Precautionary principle. Precautionary principle refers to the policy of adopting preventative measures to address potential risks to the public associated with certain activities, such as releasing a new AI product into the market.

[Note: This term has not been used in the report.]

Red teaming. Red teaming refers to a broad range of risk-management measures for AI systems, including testing vulnerabilities, identifying mitigation measures, providing feedback, and so on, typically performed by internally appointed groups of individuals acting in an adversarial role.

Safety. AI safety is a field concerned with preventing harmful consequences resulting from the use of artificial intelligence systems.

Security. AI security refers to the protection of AI systems and the data they handle from various threats and vulnerabilities. This includes safeguarding AI models, algorithms, and the infrastructure they operate on from unauthorized access, manipulation, or misuse.

Sociotechnical. Sociotechnical theory of analysis emphasizes that the social and technical aspects of a system should be treated as interdependent and not be considered in isolation from one another.

Surrogate end point. A surrogate endpoint, in the context of a clinical trial, refers to an outcome that offers indirect evidence of another desired clinical outcome. (For example, reduction in blood pressure can be a surrogate end point evidencing reduction in the risk of having a stroke.)

Supply chain. In the context of AI, a supply chain refers to the chain of actors involved in different phases of AI development such as training, testing, deployment, and integration into other systems.

Upstream (in foundation model supply chain). In a supply chain involving a particular foundation model, an upstream product is one that feeds into the foundation model in question.

Appendices

Appendix 1: How AI Is Regulated Today

As mentioned above, AI systems are used throughout society and already fall within the purview of existing laws / agency jurisdiction. The FTC, the Consumer Financial Protection Bureau (CFPB), the U.S. Department of Justice (DOJ), and the U.S. Equal Employment Opportunity Commission (EEOC) all enforce **existing discrimination laws** in housing, employment, financial services, and so on.⁷¹ What follows is a brief summary of the ways AI is currently regulated under existing enforcement structures.

The **Federal Trade Commission** has broad authority to protect consumers under Section 5(a) of the FTC Act, which “prohibits unfair or deceptive acts or practices in or affecting commerce.” Recently the commission has warned companies that they are accountable for misleading

⁷¹ Rohit Chopra et al., “Joint Statement on Enforcement Efforts against Discrimination And Bias in Automated Systems,” Federal Trade Commission, April 25, 2023, https://www.ftc.gov/system/files/ftc_gov/pdf/EEOC-CRT-FTC-CFPB-AI-Joint-Statement%28final%29.pdf.

claims related to AI.⁷² In an article on its blog, the FTC cautions companies to “consider at the design stage and thereafter the reasonably foreseeable . . . ways it could be misused for fraud or cause other harm. Then ask . . . whether such risks are high enough that you shouldn’t offer the product at all.”⁷³ The FTC has brought actions against businesses that sold or distributed potentially harmful technology when the business had not taken reasonable measures to prevent injury to consumers.⁷⁴ The FTC has also forced companies to delete algorithms when they are trained with data that was collected illegally.⁷⁵

The **Department of Justice’s Civil Rights Division** enforces constitutional provisions and federal law prohibiting discrimination across many facets of society, including in education, the criminal justice system, employment, housing, lending, and voting. In June 2022, the DOJ settled its lawsuit against Meta. The complaint alleged that Meta developed algorithms that enabled advertisers to target their housing ads based on protected characteristics under the Fair Housing Act. As part of the settlement, Meta is required to develop a new ad-delivery algorithm that addresses “racial and other disparities.”⁷⁶

The **Equal Employment Opportunity Commission** enforces federal laws that make it illegal for an employer to discriminate against an applicant or employee due to a person’s race, color, religion, sex, national origin, age, disability, or genetic information. In May 2023, the agency released a technical assistance document that focused on averting discrimination against job seekers and existing workers.⁷⁷

The **Consumer Financial Protection Bureau** oversees financial products. In May 2022, the CFPB cautioned that if credit decision technology is “too complex, opaque, or novel” to explain adverse credit decisions, companies cannot use the complexity as a defense against Equal Credit Opportunity Act violations.⁷⁸ In August 2022, the CFPB issued an interpretive rule stating

⁷² Michael Atleson, “Keep Your AI Claims in Check,” Federal Trade Commission, February 27, 2023, <https://www.ftc.gov/business-guidance/blog/2023/02/keep-your-ai-claims-check>.

⁷³ Michael Atleson, “Chatbots, Deepfakes, and Voice Clones: AI Deception for Sale,” Federal Trade Commission, March 20, 2023, <https://www.ftc.gov/business-guidance/blog/2023/03/chatbots-deepfakes-voice-clones-ai-deception-sale>.

⁷⁴ Ibid.

⁷⁵ Kate Kaye, “FTC Case against Weight Watchers Means Death for Algorithms,” Protocol, March 14, 2022, <https://web.archive.org/web/20240114131137/https://www.protocol.com/policy/ftc-algorithm-destroy-data-privacy>.

⁷⁶ Office of Public Affairs, “Justice Department Secures Groundbreaking Settlement Agreement with Meta Platforms, Formerly Known as Facebook, to Resolve Allegations of Discriminatory Advertising,” press release, U.S. Department of Justice, June 21, 2022, <https://www.justice.gov/opa/pr/justice-department-secures-groundbreaking-settlement-agreement-meta-platforms-formerly-known>.

⁷⁷ U.S. Equal Employment Opportunity Commission, “Select Issues: Assessing Adverse Impact in Software, Algorithms, and Artificial Intelligence Used in Employment Selection Procedures Under Title VII of the Civil Rights Act of 1964,” May 18, 2023, <https://www.eeoc.gov/laws/guidance/select-issues-assessing-adverse-impact-software-algorithms-and-artificial>.

⁷⁸ Consumer Financial Protection Bureau, “CFPB Acts to Protect the Public from Black-Box Credit Models Using Complex Algorithms,” May 26, 2022, <https://www.consumerfinance.gov/about-us/newsroom/cfpb-acts-to-protect-the-public-from-black-box-credit-models-using-complex-algorithms>.

that when digital marketers engage in identifying prospective customers, or in placing content to influence consumer behavior, they are generally considered service providers under the Consumer Financial Protection Act. If their actions violate federal consumer financial protection laws—for example, if companies employ an algorithm for targeted marketing—they can be held responsible.⁷⁹

The **Food and Drug Administration** regulates some AI systems through its Software as a Medical Device (SaMD) process. Software products undergo a different review process, which classes devices based on associated health risks and imposes increasingly stringent requirements that range from labeling and registration at the low end to premarket approval and clinical studies at the high end. Given the dynamic nature of AI systems that continue “learning” after approval is granted, the change control plan process is particularly salient to how the FDA regulates AI and is subject to ongoing debate.⁸⁰

Software vendors have traditionally been shielded from **liability for harm** to end users through warranty disclaimers and contractual limitations of liability; however, this may be shifting.⁸¹ The discrimination laws described above, which are increasingly being used to hold algorithms accountable, are one example of this shift.

Similarly, as part of the litigation following a large-scale Marriott data breach, a US district judge found that Marriott’s information technology service provider had a **duty of care** to Marriott’s customers to prevent a data breach.⁸² Additionally, product liability law is governed by states

⁷⁹ Consumer Financial Protection Bureau, “CFPB Warns that Digital Marketing Providers Must Comply with Federal Consumer Finance Protections,” August 10, 2022, <https://www.consumerfinance.gov/about-us/newsroom/cfpb-warns-that-digital-marketing-providers-must-comply-with-federal-consumer-finance-protections>.

⁸⁰ See Center for Devices and Radiological Health, “Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices,” Food and Drug Administration, April 22, 2024, <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-ai-ml-enabled-medical-devices>; Center for Devices and Radiological Health, “Predetermined Change Control Plans for Machine Learning-Enabled Medical Devices: Guiding Principles,” Food and Drug Administration, October 24, 2023, <https://www.fda.gov/medical-devices/software-medical-device-samd/predetermined-change-control-plans-machine-learning-enabled-medical-devices-guiding-principles>; and Stein and Dunlop, *Safe before Sale*. Numerous proposals for how the SaMD process could be adapted for AI have been put forward. See for example Eric Wu et al., “How Medical AI Devices Are Evaluated: Limitations and Recommendations from an Analysis of FDA Approvals,” *Nature Medicine* 27, no. 4 (April 2021): 582–84, <https://doi.org/10.1038/s41591-021-01312-x>; Stan Benjamens, Pranavsingh Dhunoo, and Bertalan Meskó, “The State of Artificial Intelligence-Based FDA-Approved Medical Devices and Algorithms: An Online Database,” *Npj Digital Medicine* 3, no. 1 (September 11, 2020): 1–8, <https://doi.org/10.1038/s41746-020-00324-0>; and Phoebe Clark, Jayne Kim, and Yindalon Aphinyanaphongs, “Marketing and US Food and Drug Administration Clearance of Artificial Intelligence and Machine Learning Enabled Software in and as Medical Devices: A Systematic Review,” *JAMA Network Open* 6, no. 7 (July 5, 2023): e2321792, <https://doi.org/10.1001/jamanetworkopen.2023.21792>.

⁸¹ Jey Kumarasamy and Brenda Leong, “Third-Party Liability and Product Liability for AI Systems,” International Association of Privacy Professionals (IAPP), July 26, 2023, <https://iapp.org/news/a/third-party-liability-and-product-liability-for-ai-systems>.

⁸² *Marriott v. Maryland*, October 27, 2020, https://www.govinfo.gov/content/pkg/USCOURTS-mdd-8_19-md-02879/pdf/USCOURTS-mdd-8_19-md-02879-8.pdf.

and may vary in its applicability to AI systems. (New York has a “failure to warn” category, for example, whereas other states don’t.)

The National Institute for Standards and Technology (NIST) is part of the U.S. Department of Commerce and is responsible for “creating critical measurement solutions and promoting equitable standards.”⁸³ NIST works toward the development of internet and cybersecurity standards with international standards bodies and stakeholders.⁸⁴ In response to Executive Order 13859, NIST issued a “a plan for Federal engagement in the development of technical standards and related tools in support of reliable, robust, and trustworthy systems that use AI technologies.”⁸⁵ Later, Congress passed Division E of the National Defense Authorization Act for Fiscal Year 2021; section 5301 directed NIST to create the AI Risk Management Framework (RMF).⁸⁶ Version 1 of the RMF was published in January 2023 and has since been referenced in executive orders and legislative proposals.⁸⁷ In November 2023, NIST launched the U.S. AI Safety Institute to evaluate known and emerging risks of foundation models.⁸⁸ As standards come into place, legislators in Congress or in state legislatures can reference these standards, moving them from a voluntary compliance tool to a mandate.⁸⁹ In this way, NIST is a key actor in any AI regulatory regime, though its ongoing funding challenges raise concerns about the risk of regulatory capture.⁹⁰

Congress and the executive branch oversee the process of **government use and procurement of AI systems, which are in essence premarket approval mechanisms for AI systems used by the government.** If lawmakers mandate that certain disclosures, tests, and standards accompany use and become part of the procurement process, those standards start to become a type of mandate for the private sector (at least for products and services that are applicable for government use).

In December 2020, the Trump administration’s Executive Order on Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government directed agencies to maintain an

⁸³ “About,” National Institute of Standards and Technology (NIST), July 10, 2009; updated January 11, 2022, <https://www.nist.gov/about-nist>.

⁸⁴ “AI Standards Development Activities with Federal Involvement,” NIST, August 10, 2020; updated May 2, 2024, <https://www.nist.gov/standardsgov/ai-standards-development-activities-federal-involvement>.

⁸⁵ “A Plan for Federal Engagement in Developing AI Technical Standards and Related Tools in Response to Executive Order (EO 13859),” NIST, August 10, 2019; updated April 5, 2022 <https://www.nist.gov/artificial-intelligence/plan-federal-engagement-developing-ai-technical-standards-and-related-tools>.

⁸⁶ William M. (Mac) Thornberry National Defense Authorization Act for Fiscal Year 2021, H.R. 6395, 116th Cong. (2019–2020), <https://www.congress.gov/bill/116th-congress/house-bill/6395/text>.

⁸⁷ “AI Risk Management Framework” accessed July 19, 2024, NIST, <https://www.nist.gov/itl/ai-risk-management-framework>.

⁸⁸ Paul Sandle and David Shephardson, “US to Launch Its Own AI Safety Institute,” Reuters, November 1, 2023, <https://www.reuters.com/technology/us-launch-its-own-ai-safety-institute-raimondo-2023-11-01>.

⁸⁹ See for example Federal Artificial Intelligence Risk Management Act of 2023, S. 3205, 118th Cong. (2023–2024), <https://www.congress.gov/bill/118th-congress/senate-bill/3205>.

⁹⁰ Frank Lucas et al. to Laurie Locascio, December 14, 2023, Committee on Science, Space, and Technology, Congress of the United States, House of Representatives, https://democrats-science.house.gov/imo/media/doc/2023-12-14_AISI%20scientific%20merit_final-signed.pdf.

inventory of AI use cases and to “design, develop, acquire and use AI” in a responsible manner.⁹¹ In 2022, Congress passed the Advancing American AI Act, which directed the executive branch to issue policies related to “the acquisition and use of artificial intelligence” and the “civil liberties impacts of artificial intelligence-enabled systems.”⁹²

The Biden Administration’s Executive Order on AI

The Executive Order issued under the Trump administration received only partial compliance,⁹³ leading Biden to follow up with several additional administrative actions,⁹⁴ including the Blueprint for an AI Bill of Rights⁹⁵ and a subsequent Executive Order on AI that charged agencies across government with a series of tasks tied to increasing AI adoption as well as implementing guardrails.⁹⁶ Most notably, Section 7 of Executive Order 14110 recently reminded agencies to “consider opportunities to ensure that their respective civil rights and civil liberties offices are appropriately consulted on agency decisions regarding the design, development, acquisition, and use of AI in Federal Government programs and benefits administration.”⁹⁷ Section 10 reemphasizes the provisions of the Advancing American AI Act directing the Office of Management and Budget (OMB) to develop an initial means to ensure that agency contracts for the acquisition of AI systems and services align with NIST.

New Oversight for Dual-Use Foundation Models

Section 4 of Executive Order 14110 on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence leverages the Defence Production Act to oversee foundation

⁹¹ Executive Office of the President, “Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government,” Executive Order 13960, *Federal Register*, December 3, 2020, <https://www.federalregister.gov/documents/2020/12/08/2020-27065/promoting-the-use-of-trustworthy-artificial-intelligence-in-the-federal-government&sa=D&source=docs&ust=1720637964120451&usg=AOvVaw0YsuDz76AY9MrS-2D0RO8o>.

⁹² James M. Inhofe National Defense Authorization Act for Fiscal Year 2023, H.R. 7776, 117th Cong. (2021–2022), <https://www.congress.gov/bill/117th-congress/house-bill/7776/text>.

⁹³ Ben Winters, “Compilation of Federal Govt AI Use Case Inventories,” spreadsheet, accessed July 19, 2024, https://docs.google.com/spreadsheets/d/1FH-fzqwOsifhG-rp-MB7me6W9_XZlbRFkwfQRMObfRs/edit?usp=sharing.

⁹⁴ “Administration Actions on AI,” AI.gov, accessed November 26, 2023, <https://ai.gov/actions>.

⁹⁵ “Blueprint for an AI Bill of Rights,” White House Office of Science and Technology Policy (OSTP), accessed May 10, 2024, <https://www.whitehouse.gov/ostp/ai-bill-of-rights>.

⁹⁶ White House, “Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence.”

⁹⁷ Executive Office of the President, “Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government.”

models.⁹⁸ Section 4 requires companies developing so-called dual-use foundation models⁹⁹ to provide the federal government with a description of the cybersecurity protections in place to protect model weights and the results of the foundation model's performance in "AI red-team testing" based on guidance developed by NIST.¹⁰⁰ Additionally, Section 4 includes a supply chain tracking component, requiring entities "that acquire, develop, or possess a potential large-scale computing cluster to report any such acquisition, development, or possession, including the existence and location of these clusters and the amount of total computing power available in each cluster."

Finally, states have their own regulatory oversight through their **attorney general's offices** as well as **state laws**. Notable on this front is a recent release by the California Privacy Protection Agency (CPPA) of draft automated decision-making technology (ADMT) regulations.¹⁰¹ While not final, the regulations will likely go into effect sometime in 2024 or 2025. The draft proposes requirements for businesses deploying ADMT for any "decision that produces legal or similarly significant effects concerning a consumer."¹⁰² Requirements include providing users the right to opt out of ADMT and the right to access information/disclosures about a business' use of ADMT.¹⁰³ It is wise to assume that CPPA will continue to set rules regarding the use of personal data in AI systems, and that some of the rules may create changes at the national level.

Appendix 2: How Does the FDA Work?

1. Premarket assessment

⁹⁸ Executive Office of the President, Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, *Federal Register*, October 30, 2023, <https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence>. See also "Stanford Safe, Secure, and Trustworthy AI EO 14110 Tracker," spreadsheet, accessed July 19, 2024, <https://docs.google.com/spreadsheets/d/1xOL4hkQ2pLR-IA3awLiXjPLmhleXyE5-giJ5nT-h1M/edit?usp=sharing>; and "AI Exec Order: Human-Readable Edition," Google doc, accessed July 19, 2024, <https://docs.google.com/document/d/1u-MUpA7TLO4rnrhE2rceMSjqZK2vN9ltJJ38Uh5uka4/edit>.

⁹⁹ In the executive order, a *dual-use foundation model* is defined as "an AI model that is trained on broad data; generally uses self-supervision; contains at least tens of billions of parameters; is applicable across a wide range of contexts; and that exhibits, or could be easily modified to exhibit, high levels of performance at tasks that pose a serious risk to security, national economic security, national public health or safety, or any combination of those matters, such as by: (i) substantially lowering the barrier of entry for non-experts to design, synthesize, acquire, or use chemical, biological, radiological, or nuclear (CBRN) weapons; (ii) enabling powerful offensive cyber operations through automated vulnerability discovery and exploitation against a wide range of potential targets of cyber attacks; or (iii) permitting the evasion of human control or oversight through means of deception or obfuscation."

¹⁰⁰ See Section 4.2(a)(i).

¹⁰¹ CPPA, "Draft Automated Decisionmaking Technology Regulations," December 2023, https://cppa.ca.gov/meetings/materials/20231208_item2_draft.pdf.

¹⁰² Ibid. "Decision that produces legal or similarly significant effects concerning a consumer" means a decision that results in access to, or the provision or denial of, financial or lending services, housing, insurance, education enrollment or opportunity, criminal justice, employment or independent contracting opportunities or compensation, healthcare services, or essential goods or services.

¹⁰³ State of California, "A New Landmark for Consumer Control Over Their Personal Information: CPPA Proposes Regulatory Framework for Automated Decision-Making Technology," November 27, 2023, <https://cppa.ca.gov/announcements/2023/20231127.html>.

Marketing of drugs to the public is a powerful gateway for the pharmaceutical industry: it is the primary path through which drugs enter into widespread use, and convincing doctors that a drug is the right therapy to prescribe has a significant effect on the drug's success or failure in the market. Similarly, convincing insurance companies to pay for the use of that drug offers a baseline guarantee of profitability that pharmaceutical manufacturers are strongly incentivized to seek. Finally, the level of scrutiny that drugs have traditionally undergone in the United States offers an assurance of safety and efficacy that is beneficial to the pharmaceutical industry, with wide and sweeping effects across international markets.¹⁰⁴

BOX: How premarket assessment works for the FDA¹⁰⁵

1. *Investigational new drug (IND)*. Before testing on humans, drug manufacturers must complete an IND application that includes the study design, animal test data, lead investigator's qualifications, and Institutional Review Board approval.
2. *Clinical Trials*. Studies designed to answer research questions relating to medical products are tested in humans, following the protocol outlined in the IND application. Clinical trials are typically conducted in several phases (often 3–4), progressively expanding the scale of the study and number of patients involved. At different phases, manufacturers may evaluate the safety, dosage, efficacy, side effects and adverse reactions linked to the drug, sometimes considering these in tandem with one another.¹⁰⁶
3. *New drug application (NDA)*. Once the manufacturer has completed its clinical trials and gathered evidence, it sends the evidence on to the FDA for review in the form of an NDA. The NDA includes clinical trial results, a description of the manufacturing process, facilities, quality-control procedures, product description, labeling, and (rarely) a risk evaluation and mitigation strategy (REMS). If a manufacturer needs to make a change to an approved drug, they can submit a supplemental NDA rather than begin the drug application process anew, thus abbreviating the steps involved by building on the body of evidence and documentation already compiled for the previous NDA.¹⁰⁷
4. *FDA review*. FDA staff provides a written assessment of the NDA for **safety and effectiveness** (weighing risks and benefits), appropriateness of labeling, and manufacturing process.¹⁰⁸
5. *Approval*. Approval can be granted with conditions such as the need for post-approval clinical trials or restrictions on distribution. Once approval is made, the FDA will work with the applicant to develop appropriate labeling that describes the basis for approval

¹⁰⁴ “International & Interagency Coordination,” Food and Drug Administration, updated September 18, 2018; accessed September 21, 2023, <https://www.fda.gov/food/international-interagency-coordination>.

¹⁰⁵ The information in this box is a summary of information in the CRS Report. For more details, see Congressional Research Service, “How FDA Approves Drugs and Regulates Their Safety and Effectiveness,” May 8, 2018, <https://crsreports.congress.gov/product/pdf/R/R41983>.

¹⁰⁶ Office of the Commissioner, “Step 3: Clinical Research,” Food and Drug Administration, updated January 4, 2019, <https://www.fda.gov/patients/drug-development-process/step-3-clinical-research>.

¹⁰⁷ “Approved Risk Evaluation and Mitigation Strategies (REMS),” Food and Drug Administration, updated April 28, 2021; accessed November 27, 2023, <https://www.accessdata.fda.gov/scripts/cder/remis/index.cfm?event=IndvRemsDetails.page&REMS=74>.

¹⁰⁸ Office of the Commissioner, “Step 4: FDA Drug Review,” Food and Drug Administration, April 18, 2019, <https://www.fda.gov/patients/drug-development-process/step-4-fda-drug-review>.

and how to use the drug. If a manufacturer disagrees with the FDA's decision, there are mechanisms for formal appeal.¹⁰⁹

In the pharmaceutical context, the FDA uses a benchmarking process that is both flexible and standardized in the form of **end points**: evaluative metrics tailored to each drug application that are both valid and generalizable, and reflect a particular outcome being measured. End points are established in agreement between FDA staff and drug manufacturers, and form a key part of the clinical trial process as determinants of whether the drug has successfully achieved a stated health outcome.¹¹⁰ **Surrogate end points** serve as proxies, metrics that are closely linked to more traditional end points but may enable swifter evaluation by substituting a short-term outcome for a long-term one (for example, one workshop participant referenced reduction in tumor size as an example of a surrogate end point that is clinically verifiable on a shorter timeframe than seeing a patient's cancer go into remission).¹¹¹

This approach to premarket approval allows for a tremendous amount of flexibility.¹¹² While standards and guidance shape every stage of the process, there is room for context- and drug-specific flexibility in how end points are chosen. The statutory language that guides premarket review is instructive:

§ 314.2 Purpose (FD&C Act). *The purpose of this part is to establish an efficient and thorough drug review process in order to: (a) Facilitate the approval of drugs shown to be safe and effective; and (b) ensure the disapproval of drugs not shown to be safe and effective. These regulations are also intended to establish an effective system for FDA's surveillance of marketed drugs. These regulations shall be construed in light of these objectives.*

This flexibility can be important, for example in the evaluation of drugs for rare diseases, where the FDA can approve a drug with a small clinical trial sample size. FDA staff members weigh both the challenges of obtaining a large patient sample—and the downstream effects of this, such as the likelihood of delaying needed interventions for patient populations—against the need to clearly establish evidence on the safety and efficacy of a new drug, and make decisions about when and under what conditions to relax the standard.¹¹³

Another source of flexibility has been the introduction in recent decades (starting in 1987) of **accelerated approval pathways**.¹¹⁴ These pathways—of which there are currently four—allow the approval of a new drug to be expedited in various ways, such as through the use of an intermediate clinical end point or a commitment to a shorter FDA review period. These

¹⁰⁹ Ibid.

¹¹⁰ Charlie McLeod et al., “Choosing Primary Endpoints for Clinical Trials of Health Care Interventions.”

¹¹¹ Ibid.

¹¹² Remark made by a workshop participant.

¹¹³ Congressional Research Service, “How FDA Approves Drugs and Regulates Their Safety and Effectiveness,” updated May 18, 2018, <https://crsreports.congress.gov/product/pdf/R/R41983>.

¹¹⁴ Thomas J. Hwang et al., “Association between FDA and EMA Expedited Approval Programs and Therapeutic Value of New Medicines: Retrospective Cohort Study,” *BMJ* 2020;371:m3434, <https://doi.org/10.1136/bmj.m3434>.

processes are intended to be reserved only for particularly innovative drugs that are likely to provide significant improvement over existing therapies, but in practice the majority of new drugs (60 percent in 2019) utilize one or more of these pathways. Despite this widespread use, there is limited evidence that accelerated approval pathways deliver significant additional therapeutic value to patients, and some concern that they are associated with increased risks to patient safety.¹¹⁵

In some places, premarket assessment is *already* part of the digital market, though this occurs in a piecemeal fashion:

1. **App stores.** The Apple App store and Google Play store together make up a perfect duopoly and represent gatekeepers for all applications (regardless of the level of “AI”) that are accessed by consumers via a smartphone. Apple and Google conduct premarket approval at their own behest, and set standards (SDKs) that app developers must hit. Their processes are opaque and run by organizations that are not democratically elected.
2. **Government procurement processes.** When the government is the customer purchasing an AI system, there are increasingly standardized requirements that any product or service must comply with. (See Appendix 1 for examples of existing and proposed procurement standards for pre-market scrutiny.)
3. **Software as a Medical Device (SaMD).** The FDA already oversees the release of medical devices, including some AI software. Such systems are classed according to a risk categorization and undergo differing levels of evaluation depending on the level of risk prior to deployment.¹¹⁶

2. Post-market monitoring and enforcement

The FDA also has processes to track drug and device manufacturers. It issues unique identification numbers to manufacturer facilities (both domestic and international), allowing the agency to track where drugs and devices are being manufactured and to conduct facility audits.¹¹⁷

If a product that has been approved for commercial use is determined to have caused harm to patients, the FDA has a variety of measures available to use:¹¹⁸

¹¹⁵ Cassie Frank et al., “Era of Faster FDA Drug Approval Has Also Seen Increased Black-Box Warnings and Market Withdrawals,” *Health Affairs* 33, no. 8 (August 2014): 1453–9, <https://doi.org/10.1377/hlthaff.2014.0122>.

¹¹⁶ Stein and Dunlop, *Safe before Sale*.

¹¹⁷ Office of Regulatory Affairs, “FDA’s Risk-Based Approach to Inspections,” Food and Drug Administration, January 17, 2024, <https://www.fda.gov/inspections-compliance-enforcement-and-criminal-investigations/inspection-basics/fdas-risk-based-approach-inspections>.

¹¹⁸ This section is a summary of key sections of the CRS report. For more details, see Congressional Research Service, “Enforcement of the Food, Drug, and Cosmetic Act: Select Legal Issues.”

1. **Advisory action letters.** Although not required by law, the FDA often sends letters to individuals or companies to encourage voluntary action before enforcement. There are two types of letters: *warning letters* alert entities that the agency has identified “violations of regulatory significance” and request mitigating action, warning that legal action may be imminent. *Untitled letters* are softer and are used to address violations that do not merit a warning letter, such as missing risk information on promotional materials.¹¹⁹ The FDA makes advisory letters publicly available.¹²⁰
2. **Recalls.** The FDA can recall a product that FDA considers to be in violation of the law, and does so frequently (as many as a thousand prescription drugs are recalled every year).¹²¹ Recalls can be voluntary or mandatory. Most are voluntary at the request of the FDA or initiated by the manufacturers.¹²² Manufacturers must report when they have initiated a recall, which may trigger FDA oversight. A company can ignore a recall request from the FDA, but that may risk enforcement action. In the cases listed below, the FDA can issue a mandatory recall. The process for mandatory recalls vary by product, but often begin by issuing an administrative order, which gives the product manufacturer/owner an opportunity to defend the product’s safety at an informal hearing before a presiding officer.¹²³
3. **Debarment.** The FDA has the authority to prohibit specific individuals or corporations from engaging in FDA-regulated activities (essentially ending their career) based on illegal conduct (e.g., a clinical investigator who falsifies records). Debarment can be permanent or for a set period of time. The FDA maintains a public list of debarred entities.¹²⁴
4. **Import alerts.** The FDA has the authority to issue import alerts or the ability to “refuse admission” to the US market if products “appear, from sample or otherwise” (“otherwise” may include a history of violations or a failed facility inspection) to violate the FD&C act.¹²⁵ The product is put on an import alert list to notify border officials that products should be automatically detained. Once products are detained, the owner can testify that the products are safe / abide by FDA regulations and the FDA can follow up with a determination to permit or refuse entry.¹²⁶

3. Producing information and expertise

¹¹⁹ Ibid.

¹²⁰ “Advisory Letters,” Food and Drug Administration, updated July 5, 2024, <https://www.fda.gov/inspections-compliance-enforcement-and-criminal-investigations/compliance-actions-and-activities/advisory-letters>.

¹²¹ “Recalls,” dashboard, Food and Drug Administration, accessed May 10, 2024, <https://datadashboard.fda.gov/ora/cd/recalls.htm>.

¹²² Robert H. Shmerling, “Drug Recalls Are Common,” Harvard Health Publishing, Harvard Medical School (blog), March 29, 2023, <https://www.health.harvard.edu/blog/drug-recalls-are-common-202303292907>.

¹²³ Congressional Research Service, “Enforcement of the Food, Drug, and Cosmetic Act: Select Legal Issues.”

¹²⁴ “FDA Debarment List (Drug Product Applications),” Food and Drug Administration, updated June 13, 2024, <https://www.fda.gov/inspections-compliance-enforcement-and-criminal-investigations/compliance-actions-and-activities/fda-debarment-list-drug-product-applications>.

¹²⁵ See Office of the Commissioner, “Federal Food, Drug, and Cosmetic Act (FD&C Act),” Section 801(a).

¹²⁶ Congressional Research Service, “Enforcement of the Food, Drug, and Cosmetic Act: Select Legal Issues.”

The FDA uses a range of mechanisms to elicit the generation of information about pharmaceuticals. These include:

1. **Approval processes.** The FDA collects thousands of details through the drug application process, and conducts its own independent review of that information. While no commercially confidential information is published, the FDA's validation of claims made by companies is open for public scrutiny and review.¹²⁷
2. **Clinical trials.** The FDA requires drug companies and other regulated entities to report data to the ClinicalTrials.gov database maintained by the National Institutes of Health, and engages in monitoring and enforcement to ensure compliance,¹²⁸ though this is inconsistent.¹²⁹
3. **User disclosures / labeling.** The FDA requires products in its jurisdiction to have labels, and sets rules for what these labels must look like.¹³⁰
4. **Incident reporting.** The FDA requires drug manufacturers to report serious and unexpected adverse reactions to the FDA Adverse Event Reporting System within 15 days, after which the reports are reviewed by the Office of Surveillance and Epidemiology. The Office also reviews and conducts its own studies, reviews errors in similar drugs, and follows international regulatory bodies.¹³¹
5. **Post-market studies.** The FDA may require post-market studies if it becomes aware of risks, and it can initiate enforcement action against companies if they fail to adhere to the timetable when notified by the FDA of the necessity of participating in a post-market study or clinical trial.
 - a. A 2015 Government Accountability Office (GAO) report highlighted that these post-market studies can be challenging because of the lack of incentives for clinicians and patients to participate, and that they create an additional reporting burden for medical practitioners.¹³²

A variety of stakeholders can engage in the FDA drug approval process in multiple ways:

1. **Advisory Committees.** The FDA engages advisory committees to comment on the quality of data for submitting a drug/device approval and to recommend additional studies or label changes. Advisory committee advice is nonbinding, and final decisions rest with the agency. Legally, advisory committee membership must be “fairly balanced” both in demographic diversity and experience/expertise. Most of FDA's drug advisory committees consist of physician-scientists who specialize in the drug. Industry

¹²⁷ “Drugs@FDA: FDA-Approved Drugs,” Food and Drug Administration, accessed May 10, 2024, <https://www.accessdata.fda.gov/scripts/cder/daf/index.cfm>.

¹²⁸ ClinicalTrials.Gov, accessed May 10, 2024, <https://clinicaltrials.gov>.

¹²⁹ Morten, Nicholas, and Viljoen, “Researcher Access to Social Media Data.”

¹³⁰ Congressional Research Service, “How FDA Approves Drugs and Regulates Their Safety and Effectiveness,” updated May 18, 2018, <https://crsreports.congress.gov/product/pdf/R/R41983>.

¹³¹ “Questions and Answers on FDA's Adverse Event Reporting System (FAERS),” Food and Drug Administration, updated June 4, 2018, <https://www.fda.gov/drugs/surveillance/questions-and-answers-fdas-adverse-event-reporting-system-faers>. See also “CDER Office of Surveillance and Epidemiology,” Food and Drug Administration, updated January 29, 2024, <https://www.fda.gov/about-fda/cder-offices-and-divisions/cder-office-surveillance-and-epidemiology>.

¹³² “Medical Devices: FDA Ordered Postmarket Studies to Better Understand Safety Issues, and Many Studies Are Ongoing,” U.S. Government Accountability Office, September 30, 2015, <https://www.gao.gov/products/gao-15-815>.

representatives act as individuals speaking for concerns of the industry globally, not as representatives of their employer. There are also members from consumer advocacy groups.¹³³

2. **Advisory committee meetings** (10 to 15 members) are public and open to the press. They are typically held twice a year and last two days, and expenses are covered by the FDA. Members arrive having read preparatory background materials, such as summaries regarding the safety and effectiveness of the new product.¹³⁴
3. **Patient Representatives.** “Patient Representatives” represent the closest parallel to the AI governance equivalent of involving users / data subjects in AI risk assessments. Patients and advocates are considered temporary employees who provide direct input to agency staff and engage with experts on the FDA advisory committees. Candidates are recruited and trained in how to engage in FDA activities / contribute to decisions. Applicants must have personal experience with the disease they are representing, “ability to be objective,” willingness to share their views, knowledge of treatment options, and no conflicts of interest.¹³⁵

Appendix 3: Comparison of FDA Mechanisms and AI Regulatory Proposals

| Food and Drug Administration ¹³⁶ | US AI Governance (Existing and Proposed Provisions) [numbers represent bill that include a similar or related provision] |
|--|--|
| Risk Classification | |
| Drugs in which benefits/needs are greater than risks can move through premarket approval more quickly (breakthrough therapy, accelerated approval and fast track). | AI systems are classified as high-risk ¹³⁷ based on the output/sector of decision (access to services, facial recognition, recommender systems, etc.) [6, 8, 10, 11, 12, 13, 14, 15, 16, 17]. |
| Medical devices are categorized by risk tiers (class I, class II, class III). | AI systems are classified as high-risk based on computational power / number of parameters [6, 10]. |

¹³³ “Advisory Committees: Critical to the FDA’s Product Review Process,” Food and Drug Administration, updated May 4, 2016, <https://www.fda.gov/drugs/information-consumers-and-patients-drugs/advisory-committees-critical-fdas-product-review-process>.

¹³⁴ Ibid.

¹³⁵ “About the FDA Patient Representative Program,” updated April 23, 2024, <https://www.fda.gov/patients/learn-about-fda-patient-engagement/about-fda-patient-representative-program>.

¹³⁶ Much of this column was informed by two CRS Reports: “How FDA Approves Drugs and Regulates Their Safety and Effectiveness”; and Congressional Research Services, “FDA Regulation of Medical Devices,” Food and Drug Administration, updated September 14, 2016, <https://crsreports.congress.gov/product/pdf/R/R42130>.

¹³⁷ AI governance proposals all define a set of AI systems to cover; not all are called “high-risk” in the legislative text. The companies that develop or deploy the “high-risk” AI system(s) of interest are often called “developers/deployers,” “online platforms,” “covered entities,” or something similar.

| | AI systems are classified as high-risk on market / gatekeeper power [8, 9, 34]. |
|---|--|
| Risk Assessment | |
| <p>Manufacturers and FDA assess risk throughout clinical trials and premarket approval.</p> <p>Manufacturers design their own studies/tests which are approved by the FDA.</p> <p>FDA conducts audits/inspections of clinical trials and product manufacturing sites.</p> <p>FDA accredits third-party review organizations to assist with 510(k)(medical device application) review process and inspections of foreign facilities.</p> <p>FDA engages experts and patients in approval process and post-market surveillance.</p> | <p>Deployers/developers complete risk assessments that may include safety testing, disparate impact testing, descriptions of training data, model interpretation, etc. Performed inline with standards (set with engagement through NIST or agency rulemaking) [4, 6, 9, 10, 11, 12, 13, 16, 17, 23, 24].</p> <p>Deployers/developers provide risk assessments to governing agency [6, 11, 12, 13, 16].</p> <p>Covered entity must have its online platform audited by a third party based on standards set by the governing agency [9, 16].</p> <p>Risk-assessment summaries are made publicly available [11].</p> <p>Deployers/developers self-certify to following key safety standards and mandates [8, 13].</p> <p>Governing agency issues licenses to operate AI system [10, 8 (once a size limit is reached)].</p> <p>Deployers/developers assess/document AI systems as part of government procurement process [5, 6, 14, 15].</p> <p>Deployers/developers and/or governing agency engage experts and data subjects in assessment/approval [4, 8].</p> |
| Standards | |
| <p>Manufacturers reference industry standards that have been recognized by the FDA in pre- and post-market assessments</p> <p>FDA publishes risk-mitigation guidance for nearly every class II and class III medical device category.</p> | <p>Almost every AI governance proposal includes reference to NIST standards.</p> <p>Governing agency engages stakeholders to craft voluntary codes that covered entities attest to following [9, 16].</p> |

| | |
|--|---|
| | |
| User Disclosures | |
| FDA promulgates and maintains rules for labeling and disclosures in marketing materials for products (drugs, devices, food, cosmetics). | <p>Deployers/developers alert users to the use of automated decision systems [7, 13, 21].</p> <p>Deployers/developers watermark/label AI-generated content [6, 10, 13, 18, 20, 21, 22].</p> <p>Deployers/developers publish data nutrition labels / model cards (notices on how data is used by AI systems and user-friendly explanations of key model features) [9, 10, 16, 17].</p> <p>Social media platforms disclose recommendation system weights / important features and offer users control over personal data inputs [16, 19].</p> |
| Incident Reporting | |
| <p>Manufacturers of drugs and devices must report all serious and unexpected adverse reactions to the FDA within 15 days.</p> <p>FDA oversees a process for clinicians, patients, and caregivers to voluntarily report adverse outcomes.</p> | <p>Developers/deployers of AI systems are required to report adverse safety incidents to governing entity [10].</p> <p>Governing agency maintains systems to intake complaints about online platforms (and/or mandate that online platforms set up and report on complaints) [8, 9, 16].</p> |
| Post-Market Studies | |
| FDA can require post-market studies when a drug/device is approved or as the agency becomes aware of risks. | Online platforms must provide access to data to enable academic and civil society research [9, 10, 16, 19]. |
| Supply Chain Monitoring /Export Controls | |
| <p>FDA maintains registries and unique identifiers for medical devices and manufacturing facilities.</p> <p>FDA can issue <i>import alerts</i> (“refuse admission” of products into the US market).</p> | <p>Governing agency maintains registry of high-risk AI systems [10].</p> <p>Inventories of AI use cases in government [5].</p> <p>Governing agency can limit transfer of AI</p> |

| | |
|---|--|
| FDA can mandate product recalls (and/or issue warning letters). | <p>systems to adversarial nations [10].</p> <p>Governing agency can revoke a developer/deployer's license to operate/declare the developer/deployer as noncompliant [8, 10, 13].</p> |
| Bright Lines / Banned Products | |
| FDA can prohibit medical devices deemed too risky for human use (Section 516[a] FD&C Act). | <p>AI governance proposals do not include giving a regulatory entity the ability to prohibit the development/dissemination of specific subsets of AI systems.</p> <p>There is legislation to ban specific use cases of AI systems (in nuclear weapons [27], targeted advertising [28], facial recognition [26], for children [31], in the election context [29, 30]).</p> |
| Mandates / Prohibit Conduct | |
| The FD&C Act includes statutes to mandate/prohibit certain conduct mostly with regard to the manufacturing, distribution, and marketing of food, drug, and cosmetic products. | <p><i>AI governance proposals often include language to mandate that covered entities engage in / don't engage in sets of defined conduct:</i></p> <p>Covered entities must not engage in anticompetitive conduct (e.g., self-preferencing, maintaining conflict of interest) [8, 34].</p> <p>Covered entities must provide data portability and interoperability [8, 9, 23, 24].</p> <p>Covered entities must protect personal/sensitive/covered data [1, 8, 23, 24].</p> <p>Covered entities must not process data (used in an automated decision system) in a way that discriminates [2, 8, 12, 17, 23, 24, 32].</p> <p>Covered entities offer a process for human review when decisions are made using automated decision systems (contestability) [7, 8, 10, 12, 33].</p> <p>Covered entities must not include misleading claims when marketing AI systems [1].</p> |

| Duty of Care/Tort | |
|---|--|
| Food, drugs, and cosmetics in general are subject to tort law, and if individuals are harmed they can bring a case in court. | <p>Some AI systems are likely covered by existing tort law and will not receive liability shields under Section 230 [3].</p> <p>Some proposals include text to create a duty of care for data processing / online platforms enforced by a regulator and/or state attorney general's office [8, 23, 25].</p> <p>Some proposals clarify that Section 230 does not apply to generative AI [10].</p> |
| Investigations | |
| <p>FDA can undertake investigations tied to enforcing the FD&C Act using subpoena power.</p> <p>FDA can refer cases to the Department of Justice.</p> | <p>Most AI governance proposals aimed at creating a new agency include investigative authorities to enforce laws.</p> <p>Governing agency has the authority to “study” the digital/AI market, similar to FTC 6(b) studies [8, 9].</p> |
| Funding | |
| <p>Congressional appropriations</p> <p>User fees for drug approval (negotiated with drug manufacturers that set benchmarks the agency must meet)</p> | <p>The vast majority of AI governance proposals are set up to be funded through congressional appropriations.</p> <p>Some would generate revenue through penalties paid by companies.</p> |

References for US Policies

[Enacted (or soon to be enacted) Policy]

1. Section 5 of the Federal Trade Commission Act (15 USC 45) prohibits “unfair or deceptive acts or practices in or affecting commerce”
2. Discrimination laws in DOJ's purview
3. Existing tort laws

4. NIST's work creating standards relevant to AI systems, notably AI Risk Management Framework (RMF)
5. Advancing American AI Act (passed in 2022)
6. Executive Order 14110 on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (October 2023)
7. California Consumer Privacy Act (and Draft Automated Decisionmaking Technology Regulations)

[Proposed Policy]

This list is not exhaustive; Congress and the states have introduced hundreds of laws related to AI systems (data governance).

Creating New Agencies:

8. Digital Consumer Protection Commission Act ([S.2597](#)) Sen. Warren (D-MA) & Sen. Graham (R-SC) (July 2023)
9. Digital Platform Commission Act ([S.1671](#)) Sen. Bennet (D-CO) (May 2023)
10. Blumenthal and Hawley's [Framework](#) (July 2023)

New Authorities for Existing Agencies:

11. Algorithm Accountability Act ([S.2892](#)) Sen. Wyden (D-OR) (September 2023)
12. CA AB-331 Automated Decision Tools ([amended text](#)) Bauer-Kahan (May 2023)
13. Artificial Intelligence (AI) Research, Innovation, and Accountability Act ([S.3312](#)) Sen. Thune (R-SD) & Sen. Klobuchar (D-MN) (November 2023)
14. AI Leadership To Enable Accountable Deployment Act ([S.2293](#)) Sen. Peters (D-MI) & Sen. Cornyn (R-TX) (July 2023)
15. Federal Artificial Intelligence Risk Management Act ([S.3205](#)) Sen. Moran (R-KS) & Sen. Warner (D-VA) (November 2023)
16. Digital Services Oversight and Safety Act ([H.R.6796](#)) Rep. Trahan (D-MA) (February 2022)
17. Algorithmic Justice and Online Platform Transparency Act ([S.2325](#)) Sen. Markey (D-MA) (July 2023)
18. DEEP FAKE Accountability Act ([H.R.2395](#)) Rep. Clarke (D-NY) (April 2021)
19. Platform Accountability and Transparency Act ([S.1876](#)) Sen. Coons (D-DE) & Sen. Cassidy (R-LA) (June 2023)
20. Advisory for AI-Generated Content Act ([S.2765](#)) Sen. Ricketts (R-NE) (September 2023)
21. AI Labeling Act ([S.2691](#)) Sen. Schatz (D-HI) (July 2023)
22. REAL Political Advertisements Act ([H.R.3044](#)) Rep. Clarke (D-NY) (May 2023)
23. Consumer Online Privacy Rights Act ([S.3195](#)) Sen. Cantwell (D-WA) (November 2021)
24. American Data Privacy Protection Act ([H.R.8152](#)) Rep. Pallone (D-NJ) (June 2022)
25. Data Care Act ([S.744](#)) Sen. Schatz (D-HI) (March 2023)
26. Facial Recognition and Biometric Technology Moratorium Act ([S.681](#)) Sen. Markey (D-MA) (March 2023)

27. Block Nuclear Launch by Autonomous Artificial Intelligence Act ([S.1394](#)) *Sen.Markey (D-MA)* (May 2023)
28. Banning Surveillance Advertising Act ([H.R.5534](#)) *Rep.Eshoo (D-CA)* (September 2023)
29. Protect Elections from Deceptive AI Act ([S.2770](#)) *Sen.Klobuchar (D-MN)* (September 2023)
30. Candidate Voice Fraud Prohibition Act ([H.R.4611](#)) *Rep.Espallat (D-NY)* (July 2023)
31. AI Shield for Kids Act ([S.1626](#)) *Sen.Scott (R-FL)* (May 2023)
32. Stopping Unlawful Negative Machine Impacts through National Evaluation Act ([S.5351](#)) *Sen.Portman (R-OH)* (December 2022)
33. Transparent Automated Governance Act ([S.1865](#)) *Sen.Peters (D-MI)* (June 2023)
34. American Innovation and Choice Online Act ([S.2033](#)) *Sen.Klobuchar (D-MN) & Sen.Grassley (R-IA)* (June 2023)

Further Reading

The following publications enriched our conversation; we recommend them as generative starting points for those who want to go further.

- Gianclaudio Malgieri and Frank Pasquale, “From Transparency to Justification: Toward Ex Ante Accountability for AI,” Brooklyn Law School, Legal Studies Paper No. 712, Brussels Privacy Hub Working Paper, No. 33, May 3, 2022, <https://doi.org/10.2139/ssrn.4099657>.
- Merlin Stein and Connor Dunlop, *Safe before Sale: Learnings from the FDA’s Model of Life Sciences Oversight for Foundation Models*, Ada Lovelace Institute, December 13, 2023, <https://adalovelaceinstitute.org/report/safe-before-sale>.
- Julie E. Cohen, Brenda Dvoskin, Meg Leta Jones, Paul Ohm, and Smitha Krishna Prasad, “Regulatory Monitoring in the Information Economy: Preliminary Concept Paper,” Institute for Technology Law and Policy, Georgetown Law School, October 23, 2023, <https://www.law.georgetown.edu/tech-institute/wp-content/uploads/sites/42/2023/10/10232023-Draft-Regulatory-Monitoring-Concept-Paper-1.pdf>.
- Andrew Tutt, “An FDA for Algorithms,” *Administrative Law Review* 69, no. 1 (March 15, 2016), <https://doi.org/10.2139/ssrn.2747994>.
- Daniel Carpenter, *Reputation and Power: Organizational Image and Pharmaceutical Regulation at the FDA* (Princeton: Princeton University Press, 2010), <https://press.princeton.edu/books/paperback/9780691141800/reputation-and-power>.
- Amy Kapczynski, “Dangerous Times: The FDA’s Role in Information Production, Past and Future,” *Minnesota Law Review* 102 (July 2018), <https://scholarship.law.umn.edu/mlr/130>.

Acknowledgments

Written by Anna Lenhart and Sarah Myers West, with contributions from Matt Davies and Raktima Roy.

We're grateful to those who participated in deliberation on these issues. While this report offers highlights, the group did not always arrive at consensus, and individual findings have not been, and should not be, attributed to any specific individual. Participants in the conversation included Julia Angwin, Miranda Bogen, Julie Cohen, Cynthia Conti-Cook, Matt Davies, Caitriona Fitzgerald, Ellen Goodman, Amba Kak, Vidushi Marda, Varoon Mathur, Deb Raji, Reshma Ramachandran, Joe Ross, Sandra Wachter, Sarah Myers West, and Meredith Whittaker.

We're particularly grateful to our advisory council members: Hannah Bloch-Wehba, Amy Kapczynski, Heidi Khlaaf, Chris Morten, and Frank Pasquale, and our Visiting Policy Fellow Anna Lenhart.

The deliberation was facilitated by Alix Dunn and Computer Says Maybe, with support by Alejandro Calcaño Bertorelli.