What can we learn from the FDA model for AI regulation?

10 key insights from a rapid expert deliberation on an 'FDA for AI'

1. An 'FDA for AI' is a blunt metaphor to build from. A more productive starting point would look at FDA-style regulatory interventions and how they may be targeted at different points in the AI supply chain.
2. FDA-style interventions might be better suited for certain parts of the AI supply chain than others.
3. The FDA model offers a power lesson in optimizing regulatory design for information production, rather than just product safety. This is urgently needed for AI given lack of clarity on market participants and structural opacity in AI development and deployment.
4. The lack of consensus on what counts as efficacy (rather than safety) is a powerful entry point for regulating AI. 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.
5. Pre-market approval is potentially the most powerful stage of regulatory intervention: this is where alignment between regulatory power and companies' incentives to comply reach their peak.
6. 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 is a challenge for AI given the varied provenance of AI system components, but currently characterizes the bulk of ongoing AI regulatory enforcement.
7. To have teeth, any regulatory intervention targeting the AI sector must go far beyond the current standard of penalties to meaningfully challenge some of the biggest companies in the world.
8. Greater transparency into what constitutes the market itself, and the process through which AI products are sold, will be important to AI governance. Currently the contours of what constitutes the 'AI market' are underspecified and opaque.
9. The funding model for regulatory agencies matters tremendously to its effectiveness, and can inadvertently make the regulator beholden to industry motives.
10. FDA-style documentation requirements for AI would already be a step-change from the current accountability vacuum in AI. Encouraging stronger monitoring and compliance activities within AI firms like record-keeping and documentation practices would generate organizational reflexivity as well as provide legal hooks for ex-post enforcement.

Summary overview:

This memo outlines highlights from a rapid deliberation by a group of experts who combine decades of experience studying the FDA, the pharmaceutical industry, and artificial intelligence. The group convened 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. A more detailed report of the outcomes of this discussion will be forthcoming, however this memo details a set of actionable takeaways that the conversation surfaced.

Here are key insights drawn from that conversation:

- **An 'FDA for AI' is a blunt metaphor to build from. A more productive starting point would look at FDA-style regulatory interventions and how they may be targeted at different points in the AI supply chain:**
    a. Discussions about an 'FDA for AI' often operate in a broad analogical manner, a blunt instrument for a conversation deserving of greater nuance. Using a supply chain approach to understanding AI development offers useful conceptual clarity to conversations about appropriate regulatory design rather than seeking to port the functions of a large agency whose regulatory toolbox includes many varied approaches.

- **FDA-style interventions might be better suited for certain parts of the AI supply chain than others:**
    a. The FDA's approach translates most directly at the level of the application or eventual use case where it is most tractable to validate the safety and effectiveness of an AI product.
    b. By contrast, attempting similar interventions at other stages of AI development, such as the base or 'foundation model' layer, present potentially intractable challenges like how to identify in advance the universe of possible harms using empirical evaluation. Here, other regulatory design approaches, such as financial regulation and its treatment of systemic risk, may offer more useful corollaries.
    c. At 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.
    d. It is important to clearly differentiate between the 'users' of AI applications, which are the entities procuring AI systems, and the people or communities the system is used on–the 'subjects' of AI's use. Often there is a significant power differential between 'users' and 'subjects', which regulatory interventions must also account for.

- **The FDA model offers a power lesson in optimizing regulatory design for information production, rather than just product safety. This is urgently needed for AI given lack of clarity on market participants and structural opacity in AI development and deployment.**
    a. The FDA has catalyzed and organized an entire field of expertise that has enhanced our understanding of pharmaceuticals and creating and disseminating expertise across stakeholders far beyond understanding incidents in isolation. AI is markedly opaque in contrast: mapping the ecosystem of companies and actors involved in AI development (and thus subject to any accountability or safety interventions) is a challenging task absent regulatory intervention.
    b. This information production function is particularly important for AI, a domain where the difficulty–even impossibility–of interpretability and explainability remain pressing challenges for the field and where key players in the market are incentivized against transparency. Over time, the FDA's interventions have expanded the public's understanding of how drugs work by ensuring firms invest in research and documentation to comply with a mandate to do so - prior to the existence of the agency, much of the pharmaceutical industry was largely opaque, in ways that bear similarities to the AI market.
    a. 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 multi-stakeholder 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 the public.
    b. The FDA-managed database of adverse incidents, clinical trials and guidance documentation also offers key insights for AI incident reporting (an active field of research). It may motivate shifts in the AI development process, encouraging beneficial infrastructures for increasing transparency of deployment and clearer documentation.

- **The lack of consensus on what counts as efficacy (rather than safety) is a powerful entry point for regulating AI. 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.**
    a. A lesson from the FDA is that safety and efficacy of products must be evaluated in parallel. In the context of AI, policymaking has tended to index heavily on safety

and harm and not as focused on evaluating or challenging the fundamental premise of efficacy, or concrete appraisal of risks and benefits.

b. To serve the public interest, measures of efficacy should be considered carefully so that they are not primarily or solely indexed on profit or growth, but 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 societal problems it attempts to address, and the benefits it offers, not just evaluate where it fails.

c. Efficacy evaluation could present an existential challenge to some domains and applications of AI where we currently lack the necessary methods to validate the ostensible benefits of AI usage, given widespread failures in machine learning research to reproduce the findings published in papers.

- **Pre-market approval is potentially the most powerful stage of regulatory intervention: this is where alignment between regulatory power and companies' incentives to comply reach their peak.**

    a. Past the point of market entry, the FDA retains some ability to act in the public interest, through market surveillance and recalls - but we see a significant drop in the agency's ability to act and its track record for doing so successfully.

- **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 is a challenge for AI given the varied provenance of AI system components, but currently characterizes the bulk of ongoing AI regulatory enforcement.**

    a. Looking to the FDA analogy, downstream accountability occurs through mechanisms such as recalling products after the fact, though its ability to enact these remedies is weakened once they are in commercial use. Applied to AI, this is made even more challenging given the difficulty in clearly identifying the chain of provenance for particular components of AI systems.

    b. In the context of the FDA, companies remain liable for harms caused to the public after drugs are made available for wide release, but establishing liability and then demonstrating causation in the AI context are significant barriers. Currently, the bulk of regulatory enforcement of existing law in AI occurs ex-post, and is thus subject to these challenges.

- **To have teeth, any regulatory intervention targeting the AI sector must go far beyond the current standard of penalties to meaningfully challenge some of the biggest companies in the world.**
    a. 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, and under present conditions there are no corollary gates to market entry for AI companies.
    b. The power of FDA regulation also comes from other actors in the system, from physicians to insurance companies, who can themselves refuse to recommend or cover a product if they believe it not helpful. This has acted as an important second line of defense in pharmaceuticals where the regulatory process has failed to be sufficiently rigorous, and there are corollaries in other industries such as banking and insurance. This deserves stronger development in the context of AI where the dependencies and sites of friction remain comparatively immature.

- **Greater transparency into what constitutes the market itself, and the process through which AI products are sold, will be important to AI governance. Currently the contours of what constitutes the 'AI market' are underspecified and opaque.**
    a. 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. 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?
    b. Moreover, conceptual clarity that the entity procuring the AI system is often not the same as the individual the system is used on is key, given that AI systems are frequently used by comparatively powerful entities on the less powerful, necessitating interventions that go beyond deceptive marketing and protect the interests of the public at large.

- **The funding model for regulatory agencies matters tremendously to its effectiveness, and can inadvertently make the regulator beholden to industry motives.**
    a. The FDA utilizes fees paid by industry players to fund its review process, which ensures adequate resourcing for reviews. However, under the present model the FDA must submit its budgets regularly to companies paying fees, making them

responsible to the companies it is reviewing for its accounting - this is a significant weakening of the agency's power and risks creating leverage by industry.

- **FDA-style documentation requirements for AI would already be a step-change from the current accountability vacuum in AI. Encouraging stronger monitoring and compliance activities within AI firms like record-keeping and documentation practices would generate organizational reflexivity as well as provide legal hooks for ex-post enforcement.**
    a. Introducing FDA-style functions into the AI governance process could motivate restructuring of the development practices, and potentially the operating model, of AI developers. In and of itself, this would create greater internal transparency and accountability within AI firms that would convey societal benefits, and aid the work of enforcement agencies when they need to investigate AI companies.

We're grateful to those who participated in deliberation on these issues. While this memo 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 include: Julia Angwin, Hannah Bloch-Wehba, Miranda Bogen, Alejandro Calcaño, Julie Cohen, Cynthia Conti-Cook, Matt Davies, Alix Dunn, Caitriona Fitzgerald, Ellen Goodman, Amba Kak, Amy Kapczynski, Heidy Khlaaf, Anna Lenhart, Vidushi Marda, Varoon Mathur, Chris Morten, Frank Pasquale, Deb Raji, Reshma Ramachandran, Joe Ross, Sandra Wachter, Sarah Myers West, and Meredith Whittaker.