American Principles for AI Policy

Andrew Peek[1,3] and Michael Zargham[2,3,4]

[1]Chief Executive Officer, Delphia
[2]Chief Engineer, BlockScience
[3]Trustee, Superset Trust
[4]Research Director, Metagov

March 2024

Abstract

This essay argues that for the United States to maintain global leadership in the era of artificial intelligence, it must develop a Sovereign AI landscape that reflects quintessential American values: pluralism, diversity of expression, openness to new ideas, rewarding success, and resistance to exploitation. To achieve this, policymakers should consider phenomena at three scales: the Individual (empowering data providers), Organizational (fostering a competitive marketplace of AI-powered organizations), and National (setting boundaries while supporting innovation). A key recommendation of this essay is moving beyond viewing citizens as passive participants in the coming AI economy, and instead to encourage them to actively take part in shaping the AI landscape through both their creative endeavors and their decisions regarding which organizations they share data with. To this end, the essay emphasizes the importance of striking a balance between freedoms and constraints, ensuring the sovereignty of American AI, and cultivating a dynamic AI capability that is as adaptive and resilient as America itself. Ultimately, aligning incentives between citizen knowledge creators and AI developers, along with a flexible and transparent regulatory framework, will ensure America's Sovereign AI landscape is as exceptional as the nation itself.

The recent "Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence" correctly observes that "AI reflects the principles of the people who build it, the people who use it, and the data upon which it is built." [1] Thus, it explains "[t]he reasons that America thrived in past eras of rapid change [...] are the reasons that we will succeed again in this moment" — namely, "the power of our ideals; the foundations of our society; and the creativity, diversity, and decency of our people." In other words, the United States has a great advantage in the ongoing global contest for AI supremacy, and that advantage consists of the opportunity to approach the challenges and opportunities presented by Artificial Intelligence in accordance with America's most deeply-held values — the same principles that have enabled us to lead the world, throughout our history.

What would this approach to AI entail?

A recent essay published by NVIDIA details the concept of "Sovereign AI," which "refers to a nation's capabilities to produce artificial intelligence using its own infrastructure, data, workforce and business networks."[2] The essay describes how "[n]ations around the world are already investing in [...] Sovereign AI capabilities, including ecosystem enablement and workforce development, creating the conditions for engineers, developers, scientists, entrepreneurs, creators and public sector officials to pursue their AI ambitions at home" — but it does not address the wide range of possible ways in which a given nation's Sovereign AI can reflect, or be engineered to reflect, that nation's unique character and culture.

It is precisely this last question that points the way forward for American AI: The best way to position the United States to continue leading the world through the coming age of Artificial Intelligence is to ensure that America's Sovereign AI landscape is defined by the same qualities that have historically led to our nation's success.

What are these qualities?

It would be naive to attribute America's vitality to a *laissez faire* approach to regulating markets; such an approach serves the interests of large multinational corporations, but these interests diverge substantially from those of the general public, and any regulatory regime for AI must meet the needs of America's citizens at least as well as it meets the needs of America's companies. Conversely, heavy regulation would be incommensurate with American values — extensive regulation favors incumbents, and generally supports market consolidation rather than encouraging innovation. The American economy is a knowledge economy, and knowledge economies stand to be completely transformed by the advent of AI; it is therefore critical that Americans have access to a competitive market of meaningfully differentiated AI tools.

The power of the American economy comes directly from the precise balance that it strikes between freedoms and constraints, such that it is possible to reward participation and innovation while also weeding out abuses, limiting liabilities, and facilitating access for new entrants. Our success as a nation has always come as the result of policy that maximizes the variety of healthy economic expression while protecting against the exploitation of the American public (and especially vulnerable populations) by malicious actors. In other words, our greatness, both at home and on the world stage, has always been a function of our profound commitment to a truly pluralist democracy — and the deliberate actions that we have taken as a nation to cultivate creativity and diversity from the bottom up, while curtailing predation from the top down.

How does this commitment relate to AI?

While it is too early for anybody to claim knowledge of all of the ways that the advent of AI will shape the coming decades, it is possible to draw some preliminary conclusions based on a combination of observation and reasoning about the underlying technology.

At its core, "Artificial Intelligence" is technology that automates the process of deriving meaningful patterns from datasets, such as the massive "training libraries" of written text used to parameterize Large Language Models (LLMs). As the field advances, AI systems are able to identify and replicate more — and more complex — patterns, across an increasingly wide range of kinds of data (such as natural language, audio files, and both still and moving images). AI's impact on the economy, therefore, is likely to be strongest in the realm of economic activity that depends most heavily on advanced pattern recognition: prediction.

AI's transformative potential is a consequence of the ways that such systems enable people to "get more" out of whatever data they have in their possession: more insight into the latent patterns it contains, and thus more understanding of how those patterns are likely to extend into the future, beyond the limits of the already-known. Even AI's detractors acknowledge that such systems have predictive capabilities that greatly exceed those of individual minds — hence the frequently-heard claim that LLMs are "stochastic parrots" whose alleged "thinking" consists merely of predicting what the next word in the string of text that they are generating "should" be (i.e is most likely to be, based on the patterns derived from their training data).

The concept of mining data for valuable insights is well-established; data is valuable because of the flecks of knowledge about the world that can be sifted out of it. Extending this metaphor, AI systems are essentially *refineries* that are capable of processing the raw ore that data mining surfaces into increasingly purified and concentrated — and thus increasingly valuable — forms.

Recognizing that AI's fundamental value proposition is located in its function as a data refinery leads to an unconventional conclusion — namely, that what winning the struggle for "AI supremacy" entails is developing Sovereign AI that

is better at refining data than the Sovereign AI of other nations. But what does being "better at refining data" mean, in practice?

The simple answer — *being able to use the available data to make more accurate predictions* — is partially correct, but also potentially misleading. As any one model becomes more precisely-fitted to a particular use-context, it necessarily becomes less well-fit to other potential use-contexts.[3] If America's Sovereign AI does not afford Americans the capacity for future capacities, and can only be used to do the kinds of things that have already been done, then the only thing AI will help Americans produce is stagnation.

Instead, policymakers should evaluate the capabilities of America's Sovereign AI in terms of *how well such systems are able to help Americans make* ***useful*** *predictions across a* ***variety*** *of contexts, both foreseen and unforeseen (or even currently unforeseeable)* — a definition of "intelligence" that is simultaneously more comprehensive and more sensitive to the fact that *usefulness* is always a contextually-dependent quality. To that end, America's policies must both stimulate the continued development of new AI capabilities *and* motivate citizens to contribute their data to those systems.

The American economy leads the world because its institutions are configured to optimize for its vibrancy, vitality, and dynamism — the ability to respond to changing circumstances, unforeseen challenges, and the inexorable forward march of time with continuous innovation informed by an ever-expanding range of perspectives. Unless policymakers apply the same heuristics when evaluating the intelligence of America's Sovereign AI, making our systems smarter will only accelerate an automated national decline — something that any Sovereign AI worth calling *intelligent* must instead help Americans to avert.

How can policy advance this goal?

Before maximizing the intelligence of America's Sovereign AI landscape, it is first necessary to guarantee its sovereignty. A critical — but often-overlooked — component of this sovereignty relates to the materiality of AI systems, for all such systems require hardware, sufficient power to run that hardware, and expert labor. American AI cannot be truly sovereign without accounting for all of the forms of capital and labor required to produce, operate, and maintain artificially intelligent systems: where hardware will be manufactured, where data centers will be physically located, how their electricity will be generated, and who will be providing the skilled labor required by these (and other) associated industries. The American-ness of America's Sovereign AI can only be secured by minimizing our dependence on and sensitivity to foreign actors along these axes.

Once the sovereignty of America's AI is secure, the government should seek to maximize its aggregate intelligence. The aggregate intelligence of Sovereign AI can be influenced by policy at three scales: the *Individual*, the *Organizational*, and the *National*. These three levels of scale correspond to the "Micro," "Meso," and "Macro" scales described in [4].

Behavior at the Individual scale is characterized by the relationships between organizations and the individual data producers whose data they seek to obtain, behavior at the Organizational scale is what organizations do with this data once it is in their possession, and behavior at the National scale is *directly* regulating the acceptable relationships between data producers and organizations within the jurisdiction of the nation in question. Within this configuration, patterns of Individual behavior regulate Organizational-scale *activity*, and vice versa, while the purpose of National-scale policy is to structure a field of action within which these co-regulatory processes can play out with minimal distortion. Activity at the Individual and Organizational scales also exerts regulatory influence over policy at the National scale, but over a longer time horizon, as National-scale processes typically unfold more slowly than those occurring at the Individual and Organizational scales.

Policy at the Individual scale can increase the intelligence of America's Sovereign AI by improving the quality of the data that it produces, policy at the Organizational scale can increase the intelligence of America's Sovereign AI by enhancing the health of the marketplace within which that data is produced and consumed, and policy at the National scale can increase the intelligence of America's Sovereign AI by broadening the range of perspectives that can be expressed — and thus used in forming predictions.

Overall, American AI policy should be oriented toward cultivating a breadth of options for actors at both the Individual and Organizational scales, as well as market structures that support healthy relations between them — in other words, an American Sovereign AI landscape in which a biodiverse array of different kinds of AI-powered organizations are ingesting relevant data and making predictions to inform a wide range of decisions, each in a local context.

The Individual Scale

Policy aimed at influencing Individual-scale behavior should set out to improve the quality of the available input data that America's Sovereign AI will process into predictions, by ensuring individuals receive tangible benefits for increases in the quality of the data that they provide. "Quality," in this case, refers to *how much contextually-useful information* is available in that data — roughly, what proportion of the data can credibly be treated as a meaningful "signal" rather than meaningless "noise." By definition, then, any evaluation of the quality of data depends on the specific context in which that data is to be used.

The quality of the data informing the predictions that America's Sovereign AI landscape will be able to offer can therefore be increased by expanding the set of meaningful choices that individual data contributors are given related to the collection and use of their data by AI systems, and taking actions to guarantee individuals real freedom of choice among these differentiated alternatives; the more options individual data providers have, the more clearly those providers' observed behaviors can be understood as novel signals.

Policy that empowers individual data providers will therefore improve the

capabilities of America's Sovereign AI landscape. Because their users do not have options that are meaningfully differentiated on the relevant margins, today's big tech companies simply extract data from users without engaging in the activities required to obtain informed consent, resulting in poor quality data sets and a diminished incentive for users to signal their real preferences.

Moreover, companies claim "fair use" in order to appropriate artistic and scientific works for use as training data without any accountability to the actors who originally produced them. This disrespect for cultural and intellectual production will eventually cause our society to lose its capacity to produce anything novel — which will, in turn, cause AI systems to exhaust their supply of new data to refine. It is therefore critical that government policy encourages the creators of AI systems to negotiate on a level playing field with the individuals whose data enables those systems to function. The government's simply dictating terms, however, would come at the cost of the signaletic value of market activity — and collapse the biodiversity of America's AI ecology — so *less direct* regulatory mechanisms are required. These less direct mechanisms are found at the Organizational scale.

The Organizational Scale

Rather than attempting to identify and enforce a uniform set of "fair terms" for the acquisition and use of individual providers' data, government should seek to maximize the value that such activity has *as a signal*, which is tantamount to improving the health of the economy forming around this activity — the remit of policy at the Organizational scale.

The best way of doing so is through policy that configures Organizational-scale activity into a "marketplace for acceptable behavior" by AI-powered innovators. Doing so requires ensuring that the other participants in such a marketplace, the individual data providers, have the opportunity to express their actual preferences through real choice between competing options — but also that these individual data providers have adequate visibility into the ways in which those options are meaningfully differentiated from one another. In many cases, Organizational-scale actors are insensitive to individual decisions, but other Organizational-scale actors can enable collective action in response to such insensitivity, in much the same way as labor unions emerged in response to inhumane industrial working conditions.

Government itself cannot guarantee real choice between competing options to individual data providers — although it can facilitate its emergence through policy decisions aimed at the Individual scale. Thus, policy directed at the Organizational scale must focus on empowering individuals to aggregate or organize their Individual-scale activity into Organizational-scale activity, in order to achieve a more even negotiation about what qualifies as "acceptable Organizational-scale behavior," and under which circumstances particular behaviors are or are not acceptable. Empowering individuals in this way involves providing them with adequate visibility into the differences between existing

options, and supporting the emergence of collective action organizations, such that the data individuals generate through their market behaviors reflect their actual preferences.

Government policy can accomplish this objective by supporting the establishment of independent institutions, such as Consumer Reports and the Better Business Bureau, with the mission of furnishing individuals with enough information about their options to make meaningful choices. The organization(s) filling this niche may look different from these historical antecedents (AI auditing firms such as Monitaur and O'Neil Risk Consulting & Algorithmic Auditing are already offering relevant services); what is essential is that whatever organizations emerge, they fulfill the same essential function of disseminating the information that individuals and other organizations need in order to make informed (and therefore informative) decisions between meaningfully-differentiated products and services.

The National Scale

Governance at the National scale consists of direct government regulation of the organizations that develop, operate, and supervise AI systems. Policy at this scale should aim to maximize the diversity of perspectives within America's Sovereign AI landscape, such that individuals and organizations will be able to make informed predictions. Doing so involves maximizing the amount of real choice between competing options available to the individuals providing data at the Individual and Organizational levels of scale. Thus, regulation at the National scale should be limited to only the most inviolable constraints; literally, the laws that the government asserts *all* of its constituents *must* adhere to. The purpose of National-scale AI policy should be to preserve the conditions under which the dynamic (and generative) interplay between the Individual and Organizational scales can play out, without distorting or obscuring the signals sent in the process. If National-scale policy imposes a monolithic set of values on activity at the Individual and Organizational scales, America's Sovereign AI landscape will collapse into monoculture.

Insofar as the "intelligence" of an AI system refers to that system's performance within a particular use context, and given the breadth of possible uses of AI, it can be expected that a plurality of contextually-situated AI capabilities will produce a significantly greater national AI capacity, overall, than a single AI agenda that has been centrally-planned by either the government or a corporate cartel — just as American democracy has historically provided for a diversity of institutions, both public and private, that have outperformed planned economies on sufficiently long time horizons.

Therefore, National-scale policy directing the development of America's AI landscape should be focused on setting the extreme boundaries of allowable behavior at the Organizational scale, then supporting a plurality of institutions in creating diverse forms of AI within those boundaries — similar in practice to how organizational policies set boundaries for what activities are possible at the

Individual scale. AI systems naturally resolve at the Organizational scale; to the extent that these systems are agentic, they would then be able to coordinate with each other, resulting in a networked American Sovereign AI capability that is maximally intelligent — and an accurate reflection of the plural society that it represents.

This model requires that the basic building blocks of AI systems be open-source to the greatest extent possible, but recognizes that some specific contexts (e.g., private user data, or uses specifically relevant to national security) will necessarily remain closed. The fact that these edge cases exist (and will need closer regulation), however, should not be an excuse to stymie the freedom to innovate, in general. AI may be best compared to the printing press — it has altered what information can be produced and disseminated. Wherever possible, such a capability should be widely available to artists, scientists, and journalists, whose work is needed today more than ever — even if the shape this work takes is evolving alongside the tools with which it's made.

The best predictors are ensembles — they account for multiple perspectives — and this approach will result in a proliferation of AI systems that are aligned in their adherence to core American values, but are otherwise as diverse as America itself.

Conclusion

The development of this sort of networked American Sovereign AI capability requires a stable institutional configuration that is distinctly American in its pluralism, diversity of expression, openness to new ideas, ability to reward success, and resistance to exploitation — but that nonetheless remains *governable*, despite this open-endedness.

Some aspects of this system, such as the dynamic interplay between behaviors across the Individual and Organizational scales, can only be effectively governed if they are recognized as living ecosystems. Excessive constraints often cause ecosystems to collapse — and even when they do not, they predetermine the ways in which that ecosystem must evolve, preventing it from fully manifesting its generative and creative potential.

In order to manage the transition to an AI-powered economy, we must recognize all Americans as producers of knowledge, in the form of data, and AI systems as refineries that process that raw material into intellectual, cultural, and economic capital. Aligning incentives between citizen knowledge creators and AI developers — while crafting a regulatory framework with enough flexibility and transparency to be viewed as legitimate by both groups — will ensure that America's Sovereign AI landscape is as exceptional as America itself, and that our nation's future is brightened by the lessons we have learned from our past.

References

[1] Executive Office of the President. Executive order no. 14110: "safe, secure, and trustworthy development and use of artificial intelligence". 88 Federal Register 75191, 2023. Available online at: https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence; Accessed: 2024-03-01.

[2] K. Strier. What is sovereign ai? nVIDIA Blog, 2024. Available online at: https://blogs.nvidia.com/blog/what-is-sovereign-ai/; Accessed: 2024-03-03.

[3] B. Recht. All models are wrong, but some are dangerous: Doyle's counterexample and the paradox of prediction in optimal control. *arg min* (blog), 2024. Available online at: https://www.argmin.net/p/all-models-are-wrong-but-some-are; Accessed: 2024-03-012.

[4] K. Dopfer, J. Foster, and J Potts. Micro-meso-macro. *Journal of Evolutionary Economics*, pages Springer–Verlag, 2004. Available online at: https://core.ac.uk/download/pdf/159144486.pdf; Accessed: 2024-02-29.