Honorable Gina Raimondo
Secretary
U.S. Department of Commerce
1401 Constitution Avenue N.W.
Washington, D.C. 20230

VIA ONLINE SUBMISSION

Re: Comments of Engine Advocacy in response to AI-Ready Open Data Assets Request for Information, DOC-2024-0007

Dear Secretary Raimondo:

Engine is a non-profit technology policy, research, and advocacy organization that bridges the gap between policymakers and startups. Engine works with government and a community of thousands of high-technology, growth-oriented startups across the nation to support a policy environment conducive to technology entrepreneurship. Many startups are currently developing, using, or moving towards using artificial intelligence (AI) in their products and services in diverse and innovative ways that benefit their customers and users. In order for startups to continue innovating and advancing AI technology, it is critical that they have access to abundant, high-quality, and readily usable data. We accordingly appreciate the opportunity to provide input and strongly support the Commerce Department's commitment to the responsible dissemination of public data assets for AI development.

Thriving entrepreneurship is essential to a strong and growing economy, making potential impacts on startup success a necessary consideration in many policy debates, including around AI. Repeated research has demonstrated that startups are disproportionately responsible for the innovations that drive productivity growth and economic growth, and account for virtually all net new job creation.[1]

Startups already—and will continue to—play a leading role in the development and application of AI. Startups are also likely to out-innovate and may outpace large incumbent technology companies when it comes to AI. But startups will only be able to succeed in the AI space if they have expanded access to high-quality, diverse data.

[1] *E.g.*, Ryan Decker et al., *The Role of Entrepreneurship in US Job Creation and Economic Dynamism*, 28 J. Econ. Persp. 3 (2014), http://econweb.umd.edu/~haltiwan/JEP_DHJM.pdf; John Haltiwanger et al., *High Growth Young Firms: Contribution to Job, Output and Productivity Growth*, U.S. Census Bureau Working Paper (2017), https://www.census.gov/content/dam/Census/library/working-papers/2017/adrm/carra-wp-2017-03.pdf.

Data is the lifeblood of AI applications, and developing AI can be incredibly resource-intensive.[2] Acquiring, storing, and using data to train an AI model are all expensive processes,[3] meaning AI startups—already operating with few resources[4]—are likely to have to navigate increasingly lower-margin business models, which impacts their overall competitiveness.

Larger tech companies typically possess vast data resources, either through existing services that collect proprietary data or from established partnerships with data providers. In contrast, startups often lack extensive data collection capabilities, which can impede their ability to develop and refine AI technologies. Without access to diverse, high-quality datasets, startups can struggle to train their AI models, test their algorithms, and innovate at the same pace as their larger counterparts.[5] This data disparity limits startups' competitiveness and hinders their ability to bring innovative AI solutions to market.

However, with sufficient data, startups have harnessed AI to drive innovation across nearly every industry. For example, startups are using AI to provide legal insights,[6] detect when an elderly person falls,[7] and create custom weather models for farmers.[8] This range of examples demonstrates why startups play a crucial role in AI-driven innovation, but they need access to high-quality, abundant, and diverse data to continue making significant advancements.

The Commerce Department's commitment to providing accessible and "machine understandable" data is crucial for startups to compete with incumbents. Ensuring that AI systems can more effectively interpret data without compromising quality, quantity, or integrity—through the use of knowledge graphs and standardized ontologies—lowers barriers for startups and allows them to develop competitive AI products more efficiently.

However, it is important to balance these advancements in accessibility and machine interpretability with the needs of human experts who continue to interpret and use Commerce Department data. While readily usable datasets enable startups to develop AI solutions faster, data in other forms—including raw data in obscure formats—supports innovation in other ways. To ensure that government data can satisfy the unique data requirements of different sectors and use cases, the

---

[2] *See, e.g.*, Min Jun Jung, *AI Essentials: How do neural networks work?*, Engine (July 15, 2024), https://engineadvocacyfoundation.medium.com/ai-essentials-how-do-neural-networks-work-a568cb850540.
[3] *See, e.g.*, Ivy Nguyen, *Could data costs kill your AI startup?*, VentureBeat (Nov. 10, 2018), https://venturebeat.com/2018/11/10/could-data-costs-kill-your-ai-startup/.
[4] *See, e.g.*, *the State of the Startup Ecosystem*, Engine 5-9, 16-19, (Apr. 2021) https://static1.squarespace.com/static/571681753c44d835a440c8b5/t/60819983b7f8be1a2a99972d/1619106194054/The+State+of+the+Startup+Ecosystem.pdf.
[5] *See, e.g.*, Seth Fiegerman and Matt Day, *Why AI Is So Expensive*, Bloomberg (Apr. 30, 2024), https://www.bloomberg.com/news/articles/2024-04-30/why-artificial-intelligence-is-so-expensive?embedded-checkout=true.
[6] *#StartupsEverywhere Profile: José Padilla, Founder, LegalMente AI,* Engine (Apr. 26, 2024), https://www.engine.is/news/startupseverywhere-sanantonio-tx-legalmente
[7] *#StartupsEverywhere profile: Jean Anne Booth, Founder and CEO, UnaliWear*, Engine (Apr. 30, 2021), https://www.engine.is/news/startupseverywhere-austin-tx-unaliwear
[8] *#StartupsEverywhere Profile: Carlos Gaitan, Co-Founder & CEO, Benchmark Labs*, Engine (June 28, 2024), https://www.engine.is/news/startupseverywhere-sandiego-ca-benchmark

Commerce Department should offer a range of data formats from raw to AI-ready, ensuring that specialized needs are met. Additionally, these data assets should be made available with clear licenses that allow for commercial uses, supporting an environment where startups can confidently leverage government data to drive economic growth and innovation. [9]

As the Commerce Department moves forward with AI-ready data assets, it should consider the full breadth of data that it possesses. While Bureaus like the Census Bureau and the Bureau of Economic Analysis are typically associated with data production and a wealth of information, many of the other Bureaus have data assets that are valuable for startups working with AI. For instance, a startup might use USPTO data to improve its tools helping other entrepreneurs search the prior art, making it easier to apply for novel, high-quality patents; a travel marketplace startup could use International Trade Administration data to identify tourism trends and off-seasons, helping travelers take more fulfilling vacations; and an agriculture startup can use National Oceanic and Atmospheric Administration data to identify conditions impacting pollination, helping farmers better manage their bee populations.[10] To enable a diverse range of innovation across the economy, the Department needs to ensure AI-ready data resources are developed across all its Bureaus.

Increasing access to data represents a major stride toward creating a level playing field for startups, enabling them to compete more effectively with larger tech companies, and fostering a more dynamic marketplace. Startups are adept at identifying and addressing service gaps, innovating in areas that might not be prioritized by larger entities. This enables them to provide unique solutions that address a broader range of issues and benefit a wider range of consumers, which not only enhances the overall quality and diversity of AI applications but also drives economic growth by promoting a more inclusive and varied AI ecosystem.

Engine appreciates your considering our views and the needs of U.S. startups as you continue to advance the use and dissemination of public data assets for AI development. Startups, despite having limited data resources, are a driving force in AI innovation. We strongly support your efforts to enable discovery through data and we look forward to continuing to engage with the Department of Commerce on this and other matters related to AI.

Sincerely,

Engine

700 Pennsylvania Ave S.E.
Washington, D.C. 20002
policy@engine.is

---

[9] *See, e.g., Startups and AI policy: how to mitigate risks, seize opportunities, and promote innovation*, Engine (Sep. 8, 2023), https://www.engine.is/news/category/startups-and-ai-policy-how-to-mitigate-risks-seize-opportunities-and-promote-innovation

[10] *#StartupsEverywhere: Omer Davidi, CEO and Co-Founder, BeeHero,* Engine (Nov. 6, 2020), https://www.engine.is/news/startupseverywhere-fresno-calif-beehero.