REDUCING AI'S CARBON FOOTPRINT: WHY CERTIFICATION BEATS DATA SHARING, FOR NOW

Mark Schaeffer[1]

Table of Contents

[1] Candidate for J.D., May 2026, Thomas R. Kline School of Law of Duquesne University. B.A. in Biology, 2018, Grove City College.

I. INTRODUCTION

On November 30, 2022, the technology company OpenAI released its chatbot, ChatGPT, which was capable of responding to prompts in an uncannily, human-like manner.[2] ChatGPT revolutionized the technology sector by making AI tools more accessible. AI refers broadly to computer systems that can perform tasks typically requiring human intelligence, such as recognizing patterns, making decisions, and generating language.[3] A significant subset of AI is machine learning, where algorithms learn from vast amounts of data to improve their performance over time without being explicitly programmed for every task.[4] One of the most prominent applications of machine learning today is the development of large language models ("LLMs").[5] These models are trained on massive datasets scraped from the internet including books, articles, and websites, to learn patterns in human language.[6] LLMs like ChatGPT process this data to generate human-like responses, answer questions, and simulate conversation.[7] Because they rely on huge volumes of data and complex mathematical computations, developing and deploying LLMs require enormous computing power and energy.[8]

ChatGPT attracted more than one million users in the first five days of operation alone.[9] ChatGPT's popularity prompted a rush across the business sector to either incorporate artificial intelligence ("AI") or be left behind by competitors who had already taken advantage of the new technology.[10] Since ChatGPT's release, other "big tech" companies have either released AI tools of their own or doubled down on

---

[2] Marzyeh Ghassemi et al., *ChatGPT one year on: who is using it, how and why?*, 264 NATURE 39, 39-41 (Dec. 7, 2023).
[3] DAN JURAFSKY & JAMES H. MARTIN, SPEECH AND LANGUAGE PROCESSING: AN INTRODUCTION TO NATURAL LANGUAGE PROCESSING, COMPUTATIONAL LINGUISTICS, AND SPEECH RECOGNITION 123, 220 (3d ed. draft Jan. 12, 2025), https://web.stanford.edu/~jurafsky/slp3/.
[4] *Id.*
[5] *Id.*
[6] *Id.* at 327-28.
[7] *Id.*
[8] *Id.*
[9] Ghassemi et al., *supra* note 1, at 1.
[10] Kenrick Cari, *AI 50*, FORBES (April 11, 2024, 6:30 AM), https://www.forbes.com/lists/ai50/.

their existing models.[11] Most recently, on October 4, 2024, Meta announced the release of Movie Gen, a new AI model that can generate realistic video and audio clips in response to user prompts.[12] Movie Gen was built to challenge rival tools from other leading AI tech companies like OpenAI and ElevenLabs.[13] This competition is not limited to the domestic markets: AI companies in the European Union and China have also ramped up their use and production of new AI tools.[14]

Nonetheless, despite the headlong sprint to develop new technology by nations across the globe, little focus has been given to the potential environmental impact that accompany technological advancement, particularly its effect on climate change.[15] This absence is particularly acute, as the United States Ninth Circuit Court stated, "[a]bsent some action, the destabilizing climate will bury cities, spawn life-threatening natural disasters, and jeopardize critical food and water supplies."[16]

While AI has the potential to be positively implemented for the benefit of the environment,[17] it also has enormous costs.[18] The process of training a single AI tool on human language emits more than 626,000 pounds of carbon dioxide—nearly five times the lifetime emissions of the average American car—from manufacture to junkyard.[19] Its carbon footprint has only increased due to the current AI training trends. AI developers now prioritize accuracy instead of efficiency by feeding massive

---

[11] *Id.*
[12] Katie Paul, *Meta, challenging OpenAI, announces new AI model that can generate video with sound*, REUTERS (October 7, 2024, 4:49 PM), https://www.reuters.com/technology/artificial-intelligence/meta-challenging-openai-announces-new-ai-model-that-can-generate-video-with-2024-10-04/.
[13] *Id.*
[14] Alessandro Parodi & Amir Orusov, *Governments race to regulate AI tools*, REUTERS (October 6, 2023, 7:25 AM), https://www.reuters.com/technology/governments-race-regulate-ai-tools-2023-08-22/.
[15] Patrick K. Lin, *The Cost of Training A Machine: Lighting the Way for A Climate-Aware Policy Framework That Addresses Artificial Intelligence's Carbon Footprint Problem*, 34 FORDHAM ENVTL. L. REV. 1, 6 (2023).
[16] Juliana v. U.S., 947 F.3d 1159, 1166 (9th Cir. 2020).
[17] Lin, *supra* note 8, at 6.
[18] *Id.*
[19] Emma Strubel et al., *Energy and Policy Considerations for Deep Learning in NLP*, ARXIV (June 5, 2019), https://arxiv.org/abs/1906.02243.

amounts of data to training models and trial-and-error training tactics—both of which significantly increase the carbon footprint of AI.[20]

Furthermore, the energy cost of AI does not end once the AI technology is trained: once the models are deployed in the real world for user application, they rely on inference simulate language and decisions, which calls for even more energy.[21] Moreover, the current inclination of developing "data-and power-hungry AI" may continue until more and more business sectors rely on AI to solve increasingly complex problems, exacerbating the existing environmental damage.[22] Managing the environmental consequences of AI is therefore a pressing issue.[23]

Urgently, governments must address AI's growing carbon footprint, but have so far failed to do so.[24] The lack of regulations or policies demonstrates a misplaced trust by the federal government that tech companies will voluntarily reduce their own emissions and carbon footprint.[25] However, despite tech companies' pledges to reduce carbon emission[26] there are no enforcement mechanisms or oversight to ensure they fulfill their pledges.[27] Despite the many commitments to reduce its carbon emissions, big-tech companies that employ AI, such as Google, Microsoft, Amazon, and Facebook, are still among the largest consumers of electricity in the United States.[28]

As it currently stands, federal agencies have two possible solutions they can implement to help push the future of AI in a more environmentally responsible direction: One option is promoting data sharing, which would force big tech

[20] Lin, *supra* note 8, at 6.
[21] *Id.* at 17.
[22] *Id.*
[23] *Id.*
[24] Amy L. Stein, *Artificial Intelligence and Climate Change*, 37 YALE J. ON REG. 890, 920 (2020).
[25] Lin, *supra* note 8, at 7.
[26] Stephen Shankland, *Google, Facebook, Stripe Have a $925M Plan to Capture Carbon Pollution*, CNET, (Apr. 13, 2022) https://www.cnet.com/news/google-facebook-stripe-have-a-925m-plan-to-capture-carbon-pollution/ (noting how parent companies of google and Facebook pledge nearly $1 billion to carbon capture plan).
[27] Charlotte Freitag et al., *The climate impact of ICT: A review of estimates, trends, and regulations*, 16 ENVIRONMENTAL RESEARCH LETTERS 063008 (Sept. 10, 2021). https://pubmed.ncbi.nlm.nih.gov/34553177/.
[28] *Id.* at 17.

companies to share their training data. Another option is implementing certification requirements, which would certify some AI tools as more environmentally friendly to increase consumer awareness.[29] While both of these options have potential, it is more likely that the certification requirement will actually be implemented in the future because it is easier, less intrusive, and could still have a significant impact on reducing the environmental cost of AI.

II. HISTORY

Artificial intelligence technology did not develop overnight in 2022; in fact, the algorithms that serve as the foundations for these modern tools have existed for decades.[30] Researchers have been able to develop AI tools that could beat some of the best players in the world at strategy games like Chess and Go for more than a decade.[31] In recent years, three new factors have enabled such technology to grow by leaps and bounds.[32] These new factors are: 1) the advent of massive amounts of data; 2) the ability to train the preexisting algorithms on that data; and 3) modern computing.[33] The new advances in data collection and computing have allowed the creation of powerful AI tools, which are rapidly becoming ubiquitous in modern life.[34] Beginning with the introduction of LLMs like OpenAI's ChatGPT in late 2022, similar tools were quickly followed by those from other big tech companies.[35]

[29] Stein, *supra* note 18, at 920.
[30] David R. Martinez et al., *Artificial intelligence: short history, present developments, and future outlook, final report*, MIT LINCOLN LABORATORY REPORT at 8 (2019), https://www.ll.mit.edu/r-d/publications/artificial-intelligence-short-history-present-developments-and-future-outlook.
[31] *Id.* at 17-18.
[32] *Id.* at 8.
[33] *Id.* (estimating that 90% of data in 2019 had been created since 2017).
[34] Forbes Advisor, *22 top AI statistics and trends in 2024*, FORBES (Oct. 16, 2024), https://www.forbes.com/advisor/business/ai-statistics/ (finding that 72% of businesses have adopted AI tools for at least one function).
[35] Ketmanto Wangsa et al., *A Systematic Review and Comprehensive Analysis of Pioneering AI Chatbot Models from Education to Healthcare: ChatGPT, Bard, Llama, Ernie and Grok*, 16 FUTURE INTERNET 219 (2024), https://doi.org/10.3390/fi16070219 (other models from other big tech companies include Google's Bard, Baidu's Ernie, Facebook's Llama, and Xai's Grok); *see also* Rudolph, J.; Tan, S.; Tan, S. *War of the chatbots: Bard, Bing Chat, ChatGPT, Ernie and beyond. The new AI gold rush and its impact on higher education*, J. APPL. LEARN. TEACH. (Jan. 02, 2023) 6, 364–89,

Yet all these tools and the process of training them require a lot of energy.[36] AI development begins with training the language model to operate on a large preexisting dataset that programmers and trainers use to train the system.[37] Some systems take additional feedback from users to improve.[38] By studying the provided data, the language model will begin to recognize patterns and similarities in a continuous feedback loop while it absorbs more data points.[39] The more data the system absorbs, the more its capacity will grow.[40]

A language model continues to develop even after it is released to the public as a "consumer" product.[41] Unlike traditional algorithms, which generate outputs based on fixed weights attached to predetermined input variables, LLMs continuously adjust and adapt their output weights in response to patterns identified from user interactions and other feedback.[42]

Machine learning processes drive adaptability and allow the system to analyze the outcomes selected or preferred by the user, refine its internal parameters, and iteratively optimize its responses.[43] Unlike fixed algorithms, these evolving systems are designed to improve over time, becoming more accurate and contextually aware with each new data point they process.[44] This flexibility allows AI to handle complex, non-linear problems but also introduces challenges in predictability and interpretability, as the shifting nature of these systems makes it difficult to fully understand or trace how specific outputs are derived.[45]

---

https://www.researchgate.net/publication/372689357_War_of_the_chatbots_Bard_Bing_Chat_ChatGPT_Ernie_and_beyond_The_new_AI_gold_rush_and_its_impact_on_higher_education.

[36] Tim Yarally et al., *Uncovering Energy-Efficient Practices in Deep Learning Training: Preliminary Steps Towards Green AI*, ARXIV (Mar. 24, 2023), https://arxiv.org/abs/2303.13972.

[37] Shlomit Yanisky-Ravid & Sean K. Hallisey, *Equality and Privacy by Design: A New Model of Artificial Intelligence Data Transparency Via Auditing, Certification, and Safe Harbor Regimes*, 46 FORDHAM URB. L.J. 428, 438 (2019).

[38] *Id.*

[39] *Id.* at 439.

[40] *Id.*

[41] *Id.*

[42] *Id.*

[43] *Id.*

[44] *Id.*

[45] *Id.*

Because of the constantly shifting nature and complexity of the data, it is often impossible for experts to understand how a language model arrived at a particular output.[46] Datasets are so massive and intricate that it remains unclear why the language model returned the response or produced a certain result.[47] AI language models generate their content by processing vast amounts of information collected from the internet, including websites, articles, books, and other publicly available data.[48] These models identify patterns and relationships within this data, enabling them to generate responses that mimic human language.[49] However, because the training data is so extensive and constantly evolving, tracing how a specific piece of information influenced a particular output is nearly impossible.[50] The environmental impact of these processes is significant, as the demand for electricity to power the servers, cooling systems, and infrastructure supporting AI applications grows exponentially.[51] Without adequate policies or innovations to curb this energy use, LLMs risk becoming one of the most energy-intensive industries in the modern era.[52]

User data is the most important requirement for developing any LLM.[53] These large amounts of data have made LLMs nearly ubiquitous in modern personal home technology in a short amount of time.[54] While the availability of vast datasets has driven rapid advancements in AI applications, the infrastructure required to process and store this data introduces significant environmental and economic challenges.[55]

---

[46] *Id.*
[47] *Id.*
[48] Tom B. Brown et al., *Language Models are Few-Shot Learners*, 33 ADVANCES IN NEURAL INFO. PROCESSING SYS. 1877 (2020), https://dl.acm.org/doi/pdf/10.5555/3495724.3495883.
[49] *Id.*
[50] *Id.*
[51] Karen Hao, *Training a Single AI Model Can Emit as Much Carbon as Five Cars in Their Lifetimes*, MIT TECH. REV. (June 6, 2019), https://www.technologyreview.com/2019/06/06/239031/training-a-single-ai-model-can-emit-as-much-carbon-as-five-cars-in-their-lifetimes/.
[52] *Id.*
[53] Yanisky-Ravid & Hallisey, 46 FORDHAM URB. L.J. at 439 (2019).
[54] Rudolph, *supra* note 26.
[55] Carole-Jean Wu et al., *Sustainable AI: Environmental Implications, Challenges and Opportunities*, ARXIV (Oct. 30, 2021), https://arxiv.org/abs/2111.00364.

For LLMs to make accurate inferences, a tremendous amount of processing power is necessary.[56] Particularly, storing large amounts of data requires massive data centers.[57] Each data center consumes a massive amount of energy.[58] Data center energy usage is estimated to be about two percent of the United States' total electricity usage and is expected to grow rapidly as more centers are built.[59] Data centers are one of the most energy-intensive building types, consuming ten to fifty times more energy than a typical commercial building space.[60] The largest data centers require more than 100 megawatts of power capacity—enough to power approximately 80,000 U.S. households.[61]

Nonetheless, large data centers remain a necessary byproduct of the training of these now-conventional AI tools.[62] Unsurprisingly, big tech companies like Amazon, Microsoft, Meta, and Google, which are at the forefront of the AI revolution, are also among the top ten largest data center companies.[63] As more companies seek to compete and develop their own AI tools, data centers will only continue to grow both in number and energy cost.[64] While state regulation of the environmental cost of data centers is possible,[65] federal agency regulation is the best method due to the large-scale and rapidly changing field of AI.[66] As the demand for data centers grows parallel to the expansion of AI technologies, the need for effective regulatory oversight becomes increasingly urgent.

---

[56] Lin, *supra* note 8, at 14.
[57] *Id.*
[58] U.S. DEP'T OF ENERGY, *Data Centers and Servers*, (last visited Nov. 16, 2024), https://www.energy.gov/eere/buildings/data-centers-and-servers.
[59] *Id.*
[60] *Id.*
[61] Lin, *supra* note 8, at 14; *see also 2023: These Are the World's 12 Largest Hyperscalers*, DATA CENTER KNOWLEDGE (Feb. 7, 2023), www.datacenterknowledge.com/hyperscalers/2023-these-are-the-world-s-12-largest-hyperscalers (discussing the largest data centers in the world and their electrical cost).
[62] Mary Zhang, *Top 250 Data Center Companies in the World as of 2024*, DGTL INFRA (Jan. 14, 2024), https://dgtlinfra.com/top-data-center-companies/.
[63] *Id.*
[64] *Id.*
[65] *See* Alex Engler, *A comprehensive and distributed approach to AI regulation*, THE BROOKINGS INSTITUTION (Aug. 31, 2023), https://www.brookings.edu/articles/a-comprehensive-and-distributed-approach-to-ai-regulation/.
[66] Stein, *supra* note 18, at 921.

Congress already passed legislation on January 1, 2021: the National Artificial Intelligence Initiative Act (NAIIA) was passed with bipartisan support.[67] The NAIIA provides $10 billion for federal research and development over five years.[68] NAIIA established the National Artificial Intelligence Initiative (NAII), a federal agency tasked with sustaining AI research and development and coordinating with other Federal agencies regarding AI activities.[69] This task force is responsible for investigating the feasibility of creating a national AI research cyberinfrastructure, which would provide accessible computational resources and datasets to support AI research and development.[70] The NAII aims to democratize access to AI resources, fostering innovation and diversity in the AI research community.

Multi-agency cooperation would enable the NAII to work with other Federal agencies, such as the Federal Energy Regulatory Commission (FERC) and the U.S. Department of Energy's Office of Energy Efficiency & Renewable Energy (EERE), to regulate the creation and development of AI tools.[71] This cooperation is necessary to effectively enforce potential regulations of AI tools.[72] Two potential ways in which the NAII could regulate and reduce the environmental impact of AI tools are by first, compelling data sharing between big tech companies, and second, through certification requirements.[73]

a. Mandatory Data Sharing

One potential solution to mitigate the carbon footprint of AI development is through mandatory data sharing, which could reduce the need for excessive computing resources.[74] Large data centers are the drivers of the carbon footprint of

---

[67] H.R. REP. NO. 116-617, at 1210 (2020).
[68] *Id.*
[69] Lynne Parker, *National Artificial Intelligence Initiative*, U.S. DEP'T OF COM., PATENT & TRADEMARK OFF., at 2 (Jun. 29, 2022), https://www.uspto.gov/sites/default/files/documents/National-Artificial-Intelligence-Initiative-Overview.pdf.
[70] *Id.*
[71] *Id.* at 3.
[72] *Id.* at 3.
[73] Stein, *supra* note 18, at 919.
[74] *Id.* at 920.

AI tools; thus, reducing the number of data centers necessary to train new AI tools and allow current AI tools to continue to function would keep the environmental impact in check.[75] The best means to do so would be through federal regulations. Generally, federal regulations are likely to follow trends, and data sharing per federal regulations is not new, especially in the healthcare and financial sectors.[76]

One current area of federal regulation that requires compulsory data sharing is within the healthcare sector.[77] In concert, the 21st Century Cures Act and Cares Act of 2020 enabled the CDC and other federal agencies to compel the sharing of electronic health records, clinical trial data, and administrative claims during the COVID-19 pandemic.[78] Such data sharing meant that both private and public healthcare facilities were required to keep their data in certain standardized forms and communicate it to the CDC along with other federal agencies.[79] The mandatory data-sharing policy permitted the CDC to track how the COVID-19 pandemic was affecting different communities in real-time.[80] The compulsory data sharing to promote public health in the healthcare sector is a natural analogy to compulsory data sharing in the tech sector to reduce carbon emissions.[81] Compulsory data sharing during the COVID-19 pandemic demonstrates that data sharing requirements can increase efficiency and serve the public welfare.[82]

---

[75] Stanley M. Besen, *Competition, Privacy, and Big Data*, 28 CATH. U.J.L. & TECH. 63, 77 (2020).
[76] Louis Dron et al., *Data Capture and Sharing in the COVID-19 Pandemic: A Cause for Concern*, 4 LANCET DIGIT. HEALTH 748, 748–56 (Oct. 2022), https://www.thelancet.com/action/showPdf?pii=S2589-7500%2822%2900147-9; *see also* CONSUMER FINANCIAL PROTECTION BUREAU, *Required Rulemaking on Personal Financial Data Rights* (Oct. 22, 2024) (to be codified at 12 C.F.R. pts. 1001 & 1033), https://files.consumerfinance.gov/f/documents/cfpb_personal-financial-data-rights-final-rule_2024-10.pdf.
[77] 45 C.F.R. § 170.205.
[78] Dron et al., *supra* note 76, at 748.
[79] *Id.*
[80] Dron et al., *supra* note 76.
[81] Michelle A. Williams & Gabriel Seidman, *Filling the gaps in U.S. health data*, HARVARD PUBLIC HEALTH (January 17, 2024) https://harvardpublichealth.org/policy-practice/the-u-s-public-health-data-system-is-weak-heres-how-we-fix-it/.
[82] *Id.*

Compulsory data sharing became vital during the COVID-19 pandemic.[83] The pandemic only heightened calls for increased data sharing to combat the risks of future pandemics and promote public health.[84] In the early days of the COVID-19 pandemic, public health officials were focused on addressing the crisis.[85] However, concerns over health data privacy created a barrier to decision-making.[86] The need for more data to inform better decisions was hindered by these privacy issues.[87] Advocates for greater sharing of public health data with agencies further highlight these problems.[88] Such advocates have pushed for state and local agencies to ensure that all health data is collected and stored in ways that make it easily transferable.[89] These efforts have also included making sure that privacy laws are manageable on the communication of vital health data.[90] Privacy laws in America are complicated, piecemeal, and often operate at both state and federal levels.[91] Greater synthesizing of the current data privacy laws could simplify the ability to share data in both the healthcare arena and among big tech companies as interest in AI grows.[92]

There are additional federal regulations that mandate data sharing in the financial sector.[93] The Consumer Financial Protection Bureau issued a requirement under Required Rulemaking on Personal Financial Data Rights (the "Requirement") on October 22, 2024, which mandated all financial institutions to share customers' data with other financial establishments at the request of the consumer.[94] Data

[83] Francis Collins, *Statement on Final NIH Policy for Data Management and Sharing*, NAT'L INSTS. OF HEALTH (Oct. 29, 2020), https://www.nih.gov/about-nih/who-we-are/nih-director/statements/statement-final-nih-policy-data-management-sharing.
[84] Cason Schmit, Brian N. Larson & Hye-Chung Kum, *Data Privacy in the Time of Plague*, 21 YALE J. HEALTH POL'Y L. & ETHICS 152 (Aug. 2022), https://scholarship.law.tamu.edu/facscholar/1661 at 156.
[85] *Id.*
[86] *Id.*
[87] *Id.*
[88] Williams & Seidman, *supra* note 81.
[89] *Id.*
[90] *Id.*
[91] Schmit et al, *supra* note 84 (explaining that there is no blanket privacy law in America and that different states have adopted different laws that cover some kinds of personal data and not others).
[92] Williams & Seidman, *supra* note 81.
[93] CONSUMER FINANCIAL PROTECTION BUREAU, *supra* note 76.
[94] *Id.*

sharing of this nature makes it easier for consumers to switch banks and for new companies to break into the banking market.[95] The Requirement allows customers to switch from established companies to newer ones while keeping their financial information for convenience.[96] Large financial institutions collect large amounts of data on their customers, allowing them to provide better services and products.[97] If such companies were allowed to hoard their data, it would prevent new companies from breaking into the market; failure to data share prohibits products and services from competing with the existing data-driven services and products of established large financial institutions.[98] By allowing customers to compel their banks to share data, new companies also benefit from the large data sets amassed by larger institutions.[99]

The data sharing requirements from the Consumer Financial Protection Bureau provide another compelling analogy for compulsory regulations among big tech companies aimed at reducing the energy costs of large data centers. The data sharing requirements seeks to grant consumers greater control over their data and foster increased competition within the financial sector.[100] While the goals of these data sharing requirements differ from those of regulating AI tool creation, which mandates data sharing to mitigate environmental impacts, the regulatory mechanisms operate in a similar way to those intended to reduce the carbon footprint of AI tools.[101] Nevertheless, the financial data sharing mechanics operate exactly the same as data sharing to reduce AI tools' carbon footprint by compelling private companies to share their data with each other.[102] Moreover, consumer control is emphasized in the proposed framework, which illustrates how the federal

---

[95] MORGAN LEWIS & BOCKIUS LLP, *CFPB Issues Final Rule on Personal Financial Data Rights* (Oct. 22, 2024), https://www.morganlewis.com/blogs/finreg/2024/10/cfpb-issues-final-rule-on-personal-financial-data-rights.
[96] *Id.*
[97] *Id.*
[98] *Id.*
[99] *Id.*
[100] *Id.*
[101] Stein, *supra* note 18, at 921.
[102] *Id.*

government could regulate and reduce the energy cost associated with training AI.[103] This example also highlights the broader benefits that data sharing could have regardless of the industry.[104]

Data sharing would not only significantly reduce the environmental impact of AI but also enhance competition and serve as an antitrust.[105] The antitrust benefits would assist in building momentum among the federal agencies to promote compulsory data sharing.[106] Exclusive control over large data centers makes it easier to exclude new competitors from emerging markets.[107] While it would obviously be simpler to provide incentives for companies to share data willingly, it may only sometimes be feasible due to the intense, limited competitive space and significant benefits gained by excluding new competitors.[108] Therefore, mandating data sharing as an antitrust measure could be a secondary benefit in addition to any environmental protection, making it easier for federal agencies to implement regulation in this area.[109]

i. Proposed Regulatory Body

An additional benefit of data sharing is that it puts data in the hands of the consumers.[110] A report by the Market Structure and Antitrust subcommittee has suggested that Congress should create a data regulator.[111] This proposed federal regulatory body, referred to as the Digital Authority, would have the power to compel data sharing for antitrust reasons.[112] Furthermore, the Digital Authority could set up

---

[103] Hossein Rahnama & Alex Pentland, *The New Rules of Data Privacy*, HARV. BUS. REV. (Feb. 25, 2022), https://hbr.org/2022/02/the-new-rules-of-data-privacy?form=MG0AV3.
[104] Besen, *supra* note 75, at 77.
[105] *Id.*
[106] *Id.*
[107] *Id.*
[108] *Id.* (drawing an analogy to the telecommunications industry that was compelled to require intercommunication for new competitors and that "firms with large amounts of data are also likely to be unwilling to share their data with their smaller competitors).
[109] *Id.*
[110] MARKET STRUCTURE & ANTITRUST SUBCOMM., COMM. FOR THE STUDY OF DIG. PLATFORMS, STIGLER CTR. FOR THE STUDY OF THE ECON. & THE STATE, *Report 9*, 88 (2019).
[111] *Id.*
[112] *Id.*

a mechanism that would allow consumers to choose to send their data directly from an existing big tech company to a new entrant in the field.[113]

Changing how data is managed is in line with the way data cultural perception is changing because massive amounts of data are beginning to be seen as a public good, similar to scientific knowledge.[114] The idea gaining traction is that data should not belong to a handful of companies, but instead, data should be freely shared for the common public benefit.[115] The new understanding of data could lend greater weight and momentum to the idea of compulsory data sharing.[116] Public support makes data sharing a promising possibility to curb the energy costs of AI tools.[117]

a. Certifications

A second solution would be to imitate food labeling that certifies certain products as green or environmentally friendly.[118] One such labeling system is the organic food labels organized and run by the Food & Drug Administration (FDA) and the United States Department of Agriculture (USDA).[119] Both the FDA and USDA provide ways for farms or processing facilities to sell and represent their products as organic.[120] To obtain the organic label, organic food companies are required to submit reports to a USDA agent and permit inspections of their facilities to ensure compliance.[121] Many companies go through this process in order to obtain benefits.[122] Some benefits of organic certification include greater marketing power, the ability to

---

[113] *Id.*
[114] Dana Dalrymple, *Scientific Knowledge as a Global Public Good: Contributions to Innovation and the Economy,* THE ROLE OF SCIENTIFIC AND TECHNICAL DATA AND INFORMATION IN THE PUBLIC DOMAIN: PROCEEDINGS OF A SYMPOSIUM (2003), https://www.ncbi.nlm.nih.gov/books/NBK221876.
[115] *Id.*
[116] *Id.*
[117] *Id.*
[118] *Id.*
[119] Kyle W. Lathrop, *Pre-Empting Apples with Oranges: Federal Regulation of Organic Food Labeling*, 16 J. CORP. L. 885, 887 (1991).
[120] *Organic Certification and Accreditation,* U.S. DEP'T AGRIC., https://www.ams.usda.gov/services/organic-certification (last visited Nov. 16, 2024).
[121] *Becoming a Certified Operation*, U.S. DEP'T AGRIC., https://www.ams.usda.gov/services/organic-certification/becoming-certified (last visited Nov. 16, 2024).
[122] *Id.*

sell food at higher prices, and access to funding and technical assistance that is not otherwise available.[123]

A similar certification was proposed by the Allen Institute, labeling carbon-neutral AI as "green" and non-carbon-neutral AI as "red."[124] The AI labels would operate by signaling to consumers which products are better for the environment and incentivize companies to develop energy-efficient AI.[125] Requirements for certification include algorithm, hardware, data center optimization, and pragmatic scaling.[126] Algorithm optimization is the design of optimization techniques that reduce the computational resource requirements and minimize energy consumption.[127] Hardware optimization would require AI models to be trained on more computationally efficient hardware.[128] Requiring "green" AI tools to be trained on data that optimize resource allocation, consuming as little energy as possible, could help to reduce the large carbon footprint of these data centers.[129]

Lastly, the fourth requirement for "green" AI would be to either optimize scaling or limit the number of times a LLM runs during its training process.[130] The more a LLM is trained on a data set, the more energy-costly it becomes, and the complexity increases.[131] Despite this, as AI consumes more energy, it improves less from being run through the *same* data set.[132] The result is that the most energy-intensive part of training a LLM is also the one from which the system's usefulness improves the least.[133] Having a more pragmatic approach to scaling the LLM as it gains in complexity produces a reduction in the overall energy cost of developing the

---

[123] *Benefits of Organic Certification*, U.S. DEP'T AGRIC., https://www.ams.usda.gov/services/organic-certification/benefits (last visited Nov. 16, 2024).
[124] *Id.*
[125] *Id.*
[126] Verónica Bolón-Canedo et al., *A review of green artificial intelligence: Towards a more sustainable future*, 599 NEUROCOMPUTING 128096 (Sept. 28, 2024).
[127] *Id.*
[128] *Id.*
[129] *Id.*
[130] *Id.*
[131] *Id.*
[132] Bolón-Canedo, *supra* note 120.
[133] *Id.*

AI tool.[134] The "green" labeling incentives would greatly reduce the cost of training AI tools and could be imposed similarly to the "organic" food label.[135]

Certification of AI tools as "green" would have a similar impact to organic food labeling.[136] The certification would inform consumers of the environmental costs of the products they are using while incentivizing developers of AI tools to take a more energy-efficient approach in training their LLM.[137] Both organic food labeling and certification of AI tools would have similar goals in that both grant consumers more information about products, allowing them to make environmentally beneficial choices.[138]

Although there is an element of personal safety and health in food consumption, there is also a personal health and safety element in the use of AI tools that are rapidly becoming extensions of us.[139] While both organic food labeling and AI tool certifications aim to empower consumers, there are additional considerations for AI tools that go beyond environmental concerns, particularly regarding safety and the risk of misinformation.[140] AI tools can be trained on "bad" sets of data, resulting in biased outputs, or AI tools can fall into the hands of bad actors who steal personal data and spread misinformation.[141] Using the certification, a "green" certification for an AI tool could offer not only a more environmentally friendly option but also reassurance that a Federal agency oversees the development of the LLM.[142] The

---

134 *Id.*
135 *Id.*
136 Jingwen Zhang et al., *Certification Labels for Trustworthy AI: Insights from an Empirical Study*, PROCEEDINGS OF THE 2023 CHI CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS at 1, 1–12 (2023), https://doi.org/10.1145/3593013.3593994.
137 U.S. DEP'T AGRIC., *supra* note 97.
138 *Id.*
139 Patrick Ross & Kathryn Spates, *Considering the Safety and Quality of Artificial Intelligence in Health Care*, 46 JT. COMM. J. QUAL. PATIENT SAF. 596–599 (Aug. 9, 2020), https://pmc.ncbi.nlm.nih.gov/articles/PMC7415213/pdf/main.pdf,
140 Scott Monteith et al., *Artificial intelligence and increasing misinformation*, 224 THE BRITISH JOURNAL OF PSYCHIATRY 33-35 (2024), https://www.cambridge.org/core/services/aop-cambridge-core/content/view/DCCE0EB214E3D375A3006AA69FFB210D/S0007125023001368a.pdf/artificial-intelligence-and-increasing-misinformation.pdf.
141 Katharine Miller, *Privacy in an AI Era: How Do We Protect Our Personal Information?*, STANFORD UNIVERSITY: HUMAN-CENTERED ARTIFICIAL INTELLIGENCE (Mar. 18, 2024), https://hai.stanford.edu/news/privacy-ai-era-how-do-we-protect-our-personal-information.
142 Zhang, *supra* note 130.

"green" certification can ensure that the AI model has not only met the environmental requirements but that the developers are not bad actors.[143]

III. ANALYSIS

The two methods mentioned above of regulating the environmental cost of AI tools, certification requirements, and compulsory data sharing both have great potential to curb AI's growing carbon footprint by addressing energy-intensive practices inherent to AI training and deployment.[144] In determining the most effective method, it is important to consider that each method has its own strengths and weaknesses.

Overall, the strength of compulsory data sharing is that it is more likely to reduce the carbon footprint of AI tools instantly and effectively if adequately enforced. However, this method would be much more difficult to enforce and may even run afoul of the major questions doctrine, which will be explored in further detail below.

Certifications, on the other hand, are likely to be much easier to enforce but may not decrease emissions enough to be more effective.[145] A "green" certification may even backfire and result in "greenwashing."[146] Greenwashing refers to the practice of making misleading claims about the environmental benefits of a product or service to attract environmentally conscious consumers.[147] In the context of AI, greenwashing could occur if companies falsely label their tools as environmentally friendly to improve their public image without actually making meaningful changes to reduce their carbon footprint.

In the end, it is more likely that federal agencies will introduce a "green" certification for AI tools before adopting a mandatory data-sharing requirement due to the difference in the ease of execution. Mandatory data sharing can effectively minimize the environmental cost of AI by reducing the need for redundant data

[143] Stein, *supra* note 18, at 920.
[144] Lin, *supra* note 8, at 17.
[145] Zhang, *supra* note 130.
[146] *Id.*
[147] *Id.*

processing and training efforts across different organizations. By pooling data resources, companies could limit duplicative energy usage and optimize AI training processes.[148] Such pooling could spur innovation while reducing the construction of redundant and unnecessary energy-hungry data centers.[149]

Enforcing data sharing through a federal regulation could further serve as an antitrust mechanism by limiting the power of large tech companies that have amassed substantial data resources.[150] Data sharing would enable smaller competitors to leverage existing datasets, creating a more inclusive and competitive market and preventing a few large tech companies from monopolizing data-driven advantages.[151] Allowing new entrants and smaller firms to access comparable datasets could foster competition, spur innovation, and potentially reduce the number of data centers required to support AI development.[152] Data sharing's benefit as an antitrust measure would further simplify its implementation.

Mandating data sharing, however, raises significant privacy and security issues.[153] Data is often sensitive, and sharing it across companies increases the risk of breaches and misuse.[154] A number of high-profile data breaches have only decreased trust in the security of data.[155] Developing robust mechanisms to ensure data protection and privacy compliance would be challenging, potentially stalling efforts to implement this regulation.[156]

---

[148] Abdulaziz Tabbakh et al., *Towards Sustainable AI: A Comprehensive Framework for Green AI*, Springer Journal of AI Research (2024), https://link.springer.com/article/10.1007/s43621-024-00641-4.
[149] *Id.*
[150] Denise Hearn et al., *Antitrust and Sustainability: A Landscape Analysis*, COLUMBIA LAW SCHOOL: COLUMBIA CENTER ON SUSTAINABLE INVESTMENT (2024), https://ccsi.columbia.edu/sites/default/files/content/docs/Antitrust-Sustainability-Landscape-Analysis.pdf.
[151] *Id.*
[152] *Id.*
[153] Jaspreet Bhatia & Travis D. Breaux, *Privacy Risk in Cybersecurity Data Sharing*, 2018 PROC. ACM WORKSHOP ON PRIVACY IN THE ELEC. SOC'Y 113 (2018), https://dl.acm.org/doi/pdf/10.1145/2994539.2994541.
[154] *Id.*
[155] Svetlana Abramova & Rainer Böhme, *Anatomy of a High-Profile Data Breach: Dissecting the Aftermath of a Crypto-Wallet Case*, ARXIV (2023), https://arxiv.org/abs/2308.00375.
[156] *Id.*

Even more problematic, regulations requiring companies to share proprietary data could be considered excessive government intervention in the tech industry.[157] Compulsory data sharing would likely face stiff resistance from corporations and even privacy advocates.[158] Concerns of government overreach, market disruption, and the unintended consequences of regulatory mandates would likely be difficult to assuage in the early stages of regulation of AI.

Promoting regulations that compel tech companies to share data may face significant legal challenges under the major questions doctrine. This legal principle restricts federal agencies from making decisions that exceed the historical and statutory scope of their authority without explicit congressional authorization.[159] The doctrine applies when an agency's action carries vast "economic and political significance," raising concerns about whether the agency has overstepped its legal bounds.[160]

One critical issue is the immense value associated with American data. Recent estimates place the total worth of U.S. data at approximately three trillion dollars, underscoring the substantial economic impact of any regulation that mandates data sharing among big tech companies.[161] Such a regulation would not only affect the financial structure of the tech industry but would also carry considerable political implications, as it could reshape how personal and public data are controlled and used. Therefore, the regulation would likely implicate the "economic and political significance" threshold under the second step of the major questions doctrine analysis.[162]

For an agency to enforce a mandatory data-sharing rule where the major questions doctrine is implicated, it must demonstrate a clear statutory mandate that

---

[157] Hearn et al., *supra* note 144.
[158] *Id.*
[159] Nathan Richardson, *The New Major Questions Doctrine*, 109 VA. L. REV. 923 (2023), https://virginialawreview.org/wp-content/uploads/2023/09/Deacon_Litman_Book_Revised.pdf.
[160] W. Va. v. Env't. Prot. Agency, 597 U.S. 697, 700 (2022).
[161] *See* S.O., Mai, JE. *The Ethics of Sharing: Privacy, Data, and Common Goods*, 2 DISO 28 (2023), https://link.springer.com/article/10.1007/s44206-023-00057-z.
[162] *W. Va. v. Env't. Prot. Agency*, 597 U.S. at 700.

authorizes such sweeping action.[163] Moreover, the agency must show a history of implementing similarly significant regulations—especially those involving billions of dollars—to substantiate its authority.[164] Without these elements, the regulation could face judicial scrutiny and potentially be invalidated for exceeding the agency's statutory mandate.[165]

The NAIIO, the agency established by Congress under the NAIIA,[166] would likely be unable to enact such a regulation because its mandate is limited to the following purposes:

1. Provide technical and administrative support to the Select Committee on AI (the senior interagency committee that oversees the NAII) and the National AI Initiative Advisory Committee;
2. Oversee interagency coordination of the NAII;
3. Serve as the central point of contact for technical and programmatic information exchange on activities related to the AI Initiative across federal departments and agencies, industry, academia, nonprofit organizations, professional societies, state and tribal governments, and others;
4. Conduct regular public outreach to diverse stakeholders and
5. Promote access to technologies, innovations, best practices, and expertise derived from Initiative activities to agency missions and systems across the federal government.[167]

The NAIIO's mandate limits the organization's powers to coordination and promotion rather than regulation, and certainly would not be able to regulate mandatory data sharing.

However, other agencies, such as the Federal Trade Commission (FTC), the Department of Energy (DOE), and the Environmental Protection Agency (EPA),

---

[163] Richardson, *supra* note 153.
[164] *Id.*
[165] *Id.*
[166] H.R. REP. NO. 116-617, *supra* note 67, at 1210.
[167] *Id.*

would be a different story.[168] These agencies have broad statutory mandates and have historically imposed massive regulations that have significantly affected the economy.[169] Because mandatory data sharing implicates significant financial costs and necessarily shifts the legal framework of data, it would, at the very least, trigger a major questions doctrine challenge.[170] Though mandatory data sharing is sure to reduce the carbon footprint of AI tools significantly, it remains a less attractive option to federal regulatory agencies.

On the other hand, a certification program for AI tools would be much more in line with Congress's intent in creation NAIIO of working with environmental and energy regulatory bodies.[171] NAIIO would establish "green" certification criteria, emphasizing energy efficiency, carbon-neutral practices, and transparency.[172] Compliance could be incentivized through consumer labeling, public recognition, and potential tax benefits. This approach is more politically palatable, as it encourages voluntary compliance and public engagement while minimizing regulatory burdens.[173]

Moreover, parallels already exist in other certifications, such as the "organic" food label.[174] The current certification system is minimally intrusive as it does not mandate companies to share sensitive or proprietary data but rather focuses on the output characteristics of AI tools.[175] Such an output provides flexibility and allows companies to choose their own paths to compliance.[176] Certification standards could encourage companies to adopt "best practices" in algorithm optimization, hardware

---

[168] Ann E. Ferris et al., *The Impacts of Environmental Regulation on the U.S. Economy,* OXFORD RESEARCH ENCYCLOPEDIA OF ENVIRONMENTAL SCIENCE (Sept. 26, 2017), https://oxfordre.com/environmentalscience/view/10.1093/acrefore/9780199389414.001.0001/acrefore-9780199389414-e-396.
[169] *Id.*
[170] Richardson, *supra* note 153.
[171] *Id.*
[172] Lin, *supra* note 8, at 20.
[173] Lin, *supra* note 8, at 21.
[174] *See* Zhang, *supra* 130.
[175] *See* Tabbakh, *supra* note 149.
[176] *Id.*

efficiency, and energy-conscious data management without directly disrupting business models.[177]

A green certification program can drive demand for more environmentally friendly AI products as it signals to consumers which AI tools meet specific environmental standards. Such a market-driven approach leverages consumer power to reward companies that prioritize energy efficiency, therefore creating a competitive advantage for certified products. The expected resulting public pressure and potential profitability from meeting the certification requirements will encourage tech companies to strive for greener solutions and foster a culture of sustainability within the industry.

While there is a concern that certification could lead to "greenwashing"—where companies exaggerate or misrepresent the environmental benefits of their products to meet consumer demand without making substantial changes to their operations, this arises only where there are weak standards, inadequate oversight, or a lack of transparency.[178] Greenwashing undermines the credibility and effectiveness of any certification, limiting its ability to drive genuine environmental improvement.[179] The risk can be minimized with proper oversight and a system for verifying the effectiveness of carbon capture or offset programs for AI training and applications and addressing green-washing concerns.[180]

IV. CONCLUSION

Ultimately, both mandatory data sharing and "green" certification have substantial potential to mitigate the negative environmental impact of AI technologies, but they offer different paths forward. While mandatory data sharing can potentially reduce the carbon footprint of AI tools through immediate

---

[177] *Id.*
[178] *Id.*
[179] *Id.*
[180] U.S. DEP'T OF COMMERCE: NAT'L INST. OF STANDARDS & TECH., *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile* 37 (July 2024), https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf.

optimization of data usage, it faces significant hurdles in enforcement and legal challenges, such as those posed by the major questions doctrine. The economic and political significance of requiring companies to share proprietary data also raises concerns about the feasibility of such regulations. These challenges highlight the potential difficulties in implementing such a system without clear congressional authorization or a history of similar regulations.

On the other hand, the "green" certification model offers a more politically viable and administratively feasible alternative. Certification would allow for the rapid adoption of environmentally conscious practices without imposing overly burdensome regulatory requirements on companies. Certification aligns with the current legal and market landscape by incentivizing voluntary compliance through consumer labeling, public recognition, and potential tax benefits. It allows companies to maintain flexibility while encouraging them to adopt energy-efficient practices and reduce their carbon footprints in a competitive manner. Moreover, similar certification programs, such as the "organic" food label, suggest that this model can effectively encourage positive environmental behavior without significant disruptions to current business models.

Despite concerns about the risk of "greenwashing," the certification approach provides a viable solution to the challenge of fostering a more sustainable AI industry. The key to minimizing greenwashing lies in developing robust and transparent standards, along with proper oversight to ensure compliance. With consumer demand for environmentally friendly products on the rise, the certification system could create a competitive advantage for companies prioritizing sustainability. This would reduce the environmental costs associated with AI and promote a broader cultural shift towards sustainability in the tech industry.

While data sharing remains an important long-term goal, the political, legal, and practical challenges make it less likely to be implemented in the short term. As the AI industry grows, there may be increasing public and political support for stronger regulatory measures that could address data usage and environmental concerns more comprehensively. However, the likely path forward is through incremental steps, with certification programs taking precedence due to their ease of implementation, lower political resistance, and the ability to generate immediate consumer-driven outcomes.