Global Economy and Development
at BROOKINGS

REPORT
NOV 2023

TOWARD INTERNATIONAL COOPERATION ON FOUNDATIONAL AI MODELS

AN EXPANDED ROLE FOR TRADE AGREEMENTS AND INTERNATIONAL ECONOMIC POLICY

JOSHUA P. MELTZER

Toward international cooperation on foundational AI models

An expanded role for trade agreements and international economic policy

Joshua P. Meltzer

November 2023

Acknowledgement

The Brookings Institution is a nonprofit organization based in Washington, D.C. Our mission is to conduct in-depth, nonpartisan research to improve policy and governance at local, national, and global levels. The conclusions and recommendations of any Brookings publication are solely those of its author(s), and do not reflect the views or policies of the Institution, its management, its other scholars, or the funders acknowledged below.

Brookings gratefully acknowledges the financial support of IBM and google.org for its Global Economy and Development program. Meta, PwC, McKinsey & Company, and Goldman Sachs, all mentioned in the publication, are also donors to the Institution.

Brookings recognizes that the value it provides is in its absolute commitment to quality, independence, and impact. Activities supported by its donors reflect this commitment.

Executive summary

Foundational AI presents new opportunities for social and economic flourishing, but also risks of harm

The development of artificial intelligence (AI) presents significant opportunities for economic and social flourishing. The release of foundational models such as the large language model (LLM) ChatGPT4 in early 2023 captured the world's attention, heralding a transformation in our approach to work, communication, scientific research, and diplomacy. According to Goldman Sachs, LLMs could raise global GDP by 7 percent and lift productivity growth by 1.5 percent over 10 years. McKinsey found that generative AI such as ChatGPT4 could add $2.6-$4.4 trillion each year over 60 use cases, spanning customer operations, marketing, and sales, software engineering, and R&D.[1] AI is also impacting international trade in various ways, and LLMs bolster this trend. The upsides of AI are significant and achieving them will require developing responsible and trustworthy AI. At the same time, it is critical to address the potential risk of harm not only from conventional AI but also from foundational AI models, which in many cases can either magnify existing AI risks or introduce new ones.

For example, LLMs are trained on data that encodes existing social norms, with all their biases and discrimination. LLMs create risks of information hazards by providing information that is true and can be used to create harm to others, such as how to build a bomb or commit fraud.[2] A related challenge is preventing LLMs from revealing personal information about an individual that is a risk to privacy. In other cases, LLMs will increase existing risks of harm, such as from misinformation which is already a problem with online platforms or increase the incidence and effectiveness of crime. LLMs may also introduce new risks, such as risks of exclusion where LLMs are unavailable in some languages.

International cooperation on AI is already happening in trade agreement and international economic forums

Many governments are either regulating AI or planning to do so, and the pace of regulation has increased since the release of ChatGPT4. However, regulating AI to maximize the upsides and minimize the risks of harm without stifling innovation will be challenging, particularly for a rapidly evolving technology that is still in its

[1] McKinsey, The economic potential of generative AI: The next productivity frontier. https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier#introduction

[2] N. Bostrom et al, Information Hazards: A typology of potential harms from Knowledge, Review of Contemporary Philosophy, 2011

relative infancy. Making AI work for economies and societies will require getting AI governance right. Deeper and more extensive forms of international cooperation can support domestic AI governance efforts in a number of ways. This includes by facilitating the exchange of AI governance experiences which can inform approaches to domestic AI governance; addressing externalities and extraterritorial impacts of domestic AI governance which can otherwise stifle innovation and reduce opportunities for uptake and use of AI; and finding ways to broaden access globally to the computing power and data needed to develop and train AI models.

Free trade agreements (FTAs), and more recently, digital economy agreements (DEAs) already include commitments that increase access to AI and bolster its governance. These include commitments to cross-border data flows, avoiding data localization requirements, and not requiring access to source code as a condition of market access, all subject to exception provisions that give government the policy space to also pursue other legitimate regulatory goals such as consumer protection and guarding privacy. Some FTAs and DEAs such as the New Zealand-U.K. FTA and the Digital Economy Partnership Agreement include AI-specific commitments focused on developing cooperation and alignment, including in areas such as AI standards and mutual recognition agreements.

With AI being a focus of discussions, international economic forums such as the G7 and the U.S.-EU Trade and Technology Council (TTC), the Organization for Economic Cooperation and Development (OECD), as well as the Forum for Cooperation on Artificial Intelligence (FCAI) jointly led by Brookings and the Center for European Policy Studies as a track-1.5 dialogue among government, industry, and civil society, are important for developing international cooperation in AI. Initiatives to establish international AI standards in global standards development organizations (SDOs) such as the International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) are also pivotal in developing international cooperation on AI.

But more is needed—where new trade commitments can support AI governance

These developments in FTAs, DEAs, and in international economic forums, while an important foundation, need to be developed further in order to fully address the opportunities and risks from foundational AI models such as LLMs. International economic policy for foundational AI models can use commitments in FTAs and DEAs and outcomes from international economic forums such as the G7 and TTC as mutually reinforcing opportunities for developing international cooperation on AI governance. This can happen as FTAs and DEAs elevate the output from AI-focused forums and standard-setting bodies into trade

commitments and develop new commitments as well. FCAI is another forum to explore cutting-edge AI issues.

The following table outlines key opportunities and risks from foundational AI models and how an ambitious trade policy can further develop new commitments that would help expand the opportunities of foundational AI models globally and support efforts to address AI risks, including by building on developments in forums such as the G7 and in global SDOs.

Table 1. New commitments in FTAs, DEAs and for discussion in international economic forums

Enable AI opportunity	
Increase access to AI compute and data	• Reduce barriers to hardware, data, and access to cloud computing.
Increase access to AI products and services	• Reduce barriers to AI services and AI-enabled goods.
Support opportunities to develop and use AI globally	• Commit to a dialogue and work program that identifies opportunities to cooperate on expanding AI access and use in other countries.
Manage AI risks	
Discrimination, exclusion, and toxicity	• Agree to implement appropriate privacy regulations. • Commit to internationally recognize AI ethical principles. • Develop government procurement commitments to drive responsible and trustworthy AI. • Agree to develop mutual recognition agreements related to conformity assessment and AI audits. • Include the G7 Code of Conduct for AI in trade agreements. • Commit to cooperate in developing international AI standards. • Include a TBT-style commitment to base domestic regulation on international AI standards. • Agree to share best practices around data governance.
Security and privacy	• Develop government procurement commitments to drive responsible and trustworthy AI. • Include the G7 Code of Conduct for AI in trade agreements. • Agree to implement appropriate privacy regulations. • Commit to cooperate in developing international AI standards. • Develop a TBT-style commitment to base domestic regulation on international AI standards. • Include as a trade commitment the OECD principles on government access to personal data. • Agree to share best practices around AI governance.
Misinformation	• Identify opportunities to expand cooperation on misinformation/disinformation • Include the G7 Code of Conduct for AI in trade agreements.
Explainable and interpretable results	• Commit to cooperate on the development of international AI standards. • Develop a TBT-style commitment to base domestic regulation on international AI standards. • Agree to develop mutual recognition agreements related to conformity, assessment, and AI audits. • Cooperate on the development of technical solutions.

	• Agree to share best practices around AI governance.
Measuring AI risk and accountability	• Develop a SPS-style commitment to base AI regulation on a risk assessment. • Commit to cooperate in the development of international AI standards. • Develop a TBT-style commitment to base domestic regulation on international AI standards. • Include the NIST AI RMF as a trade commitment. • Agree to share experience on AI governance. • Include the G7 Code of Conduct for AI in trade agreements.
Copyright infringement	• Agree to share developments in domestic laws and evolving approaches to foundational AI and copyright.

Introduction

The development of artificial intelligence (AI) presents significant opportunities for economic and social flourishing. The release of ChatGPT4 in early 2023 captured the world's attention, promising to change how we work, communicate, do science, and conduct diplomacy. ChatGPT4 is a large language model (LLM), which itself is a foundational AI system—one that is increasingly generalizable in that it can work across contexts and learn as it scales. Other large language models (LLMs) include Google's PaLM and Meta's LLaMA, to name a few. Foundational AI demonstrates the new opportunities as well as the risks from AI, underscoring the need for international cooperation.

This paper takes the view that the upsides of AI are significant and that achieving them will require developing responsible and trustworthy AI. Many governments are either regulating AI or planning to do so with these goals in mind, and the pace of AI policy development and regulation has increased since the release of ChatGPT4.[3] Yet, regulating AI to maximize the upsides and minimize the risks of harm without stifling innovation will be challenging, particularly for a technology that remains in its relative infancy and is fast-moving. Yet, making AI work for economies and societies will require getting AI governance right. Deeper and more extensive forms of international cooperation can help by sharing the various and different experiences with regulating AI; developing ways to address the spillovers and extraterritorial impacts of domestic AI governance; and finding ways to expand access to the data and the AI compute (the computational resources required for AI such as GPUs/TPUs and memory) needed to build and run foundational AI models consistent with the goal of responsible and trustworthy AI.

Trade agreements and more recently digital economy agreements (DEAs) already include commitments that increase access to AI and support AI governance. At the same time, AI is a focus of discussions in international economic forums such as the G7 and the U.S.-EU Trade and Technology Council (TTC). This paper focuses on how trade agreements, DEAs and key international economic forums such as the G7 and the TTC can build effective forms of international cooperation on AI governance.

Part 1 explains what a foundational AI model is, with a focus on ChatGPT4. This part also provides an overview of the impact of AI on economic opportunity and international trade, as well as its geostrategic implications, and outlines where foundational AI introduces new risks or heightens existing AI risks. Part 2 makes

[3] OECD AI Policy Observatory https://oecd.ai/en/wonk/national-policies-2

the case for why international cooperation on AI is needed to realize the opportunities of AI and build effective AI governance. This part describes how trade agreements, DEAs, and steps taken in international economic forums are already working to build international cooperation in AI. Part 3 explores how trade policy needs to be further developed to respond to the opportunities and risks from foundational AI models. Part 4 concludes.

Part 1: The opportunities and risks from foundational AI models

What are foundational AI models?

This paper focuses on foundational AI models that include large language models (LLMs) such as ChatGPT4. An LLM processes and understands natural language data such as written text, spoken words, or other forms of language input. ChatGPT4 can process visual inputs using machine learning techniques such as deep neural networks to analyze and generate human-like language based on the patterns and structures it has learned from this data.[4] LLMs are often referred to as generative AI as these models generate new content based on prompts.[5]

Foundational models such as LLMs have several key features. First is the capacity for transfer learning, where knowledge gained from training on one task, such as object recognition, can be applied to another task.[6] This means that foundational models are increasingly generalizable in that they can be used across a wide range of applications.[7] The second key element is that scaling the AI compute and training data results in significant performance improvements.[8] To put this in perspective, the computation used to train AI has scaled by a factor of 10 every year for the last 10 years. This means that each next generation of LLM will be even more powerful and impactful. Third, ever-larger datasets, exponential increases in AI compute, and the number of parameters of foundational AI models have led to new capabilities emerging as the system scales.[9] In other words, foundational AI models can develop new capabilities to perform tasks for which the AI system was not originally programmed. For example, ChatGPT4 seems to have developed in-context learning, enabling the LLM to adapt to downstream tasks by developing a description of that task.[10] Indeed, the capacity of ChatGPT4 is still being understood. Some argue that theory-of-mind (TOM)—the ability to impute unobservable mental states such as desires and beliefs to others emerged in Chat GPT3 as a byproduct of being trained to achieve other goals where TOM

[4] Definition generated by ChatGPT.
[5] ChatGPT4 Technical Report, 27 March 2023
[6] Bommasani, D.A Hudson, E. Adeli, et al., "On the opportunities and Risks of Foundation Models" https://arxiv.org/pdf/2108.07258.pdf
[7] Bommasani, D.A Hudson, E. Adeli, et al "On the opportunities and Risks of Foundation Models" https://arxiv.org/pdf/2108.07258.pdf
[8] Jacob Devline et al, BERT: Pre-training of deep bidirectional transformers for language understanding, NAACL 2019
[9] Jason Wei et al, "Emergent Abilities of Large Language Models", Transaction on Machine Learning Research, 26 October 2022
[10] Bommasani, D.A Hudson, E. Adeli, et al "On the opportunities and Risks of Foundation Models" 2108.07258.pdf (arxiv.org)

would be a benefit.[11] When it comes to ChatGPT4, some claim that elements of artificial general intelligence (AGI) may also have emerged.[12]

The social, scientific, and economic opportunities from foundational AI

Foundational AI models expand on many of the economic and social opportunities of AI. The impact of LLMs is potentially transformative given the central role of language in human culture and as the basis on which we understand the world. As Yuval Noah Harari put it recently with respect to GPT4, "In the beginning was the word. Language is the operating system of human culture. AI's new mastery of language means it can now hack and manipulate the operating system of civilization."[13] For instance, foundational AI models can write and compose music and generate images. According to an op-ed by Henry Kissinger, Eric Schmidt, and Daniel Huttenlocher, LLMs like ChatGPT4 "will redefine human knowledge, accelerate change in the fabric of our reality, and reorganize politics and society."[14]

These observations underscore the potentially wide-ranging social and economic implications of LLMs. On the economic front, LLMs could lead to rapid increases in productivity and economic growth. According to PwC's Global Artificial Intelligence Study, with accelerated development and uptake of AI, global GDP could be 14 percent or almost $16 trillion higher by 2030. According to Goldman Sachs, LLMs could raise global GDP by 7 percent and lift productivity growth by 1.5 percent over 10 years.[15] McKinsey found that generative AI such as ChatGPT4 could add $2.6-$4.4 trillion annually across the 63 use cases it analyzed, with 75 percent of that value being derived from customer operations, marketing, and sales, software engineering, and R&D.[16] Currently, large companies and large tech companies specifically have the resources—the computational capacity, data, and talent to build and train foundational AI models. However, access to foundational AI is often available via application programming interfaces (APIs) which allow further training and fine-tuning of the model for specific use-cases.

[11] Michael Kosinski, "Theory of Mind May Have Spontaneously Emerged in large Language Models", Stanford University.
[12] Sebastien Bubeck et al, Sparks of Artificial General Intelligence: Early experiments with GPT-4", arXiv:2303.12712v5, 13 April 2023.
[13] Yuval Harari, Tristan Harris, and Aza Raskin, "You can have the blue pill or the red pill, and we're out of blue pills", New York Times Guest Essay, March 23, 2023
[14] Henry Kissinger, Eric Schmidt, and Daniel Huttenlocher, "ChatGPT Heralds and Intellectual Revolution", WSJ Opinion, Feb 24, 2023
[15] Generative AI Could Raise Global GDP by 7%. https://www.goldmansachs.com/intelligence/pages/generative-ai-could-raise-global-gdp-by-7-percent.html
[16] McKinsey, The economic potential of generative AI: The next productivity frontier https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier#introduction

LLMs and foundational AI more broadly are likely to transform companies, change business models, and deeply impact jobs and the future of work.[17] This will include expanded use of robotics, product R&D, and opportunities for sales. Foundational AI should lead to more efficient manufacturing and supply chains, as well as productivity gains across services as foundational AI systems assist in information retrieval and support services delivery across education, health care, and professional services.

Foundational AI models can also drive important advances in human well-being and flourishing. For instance, AlphaFold developed by Deep Mind has predicted the structure of about 350,000 proteins—about half of all known human proteins—and is now using AI to predict how these proteins work together.[18] This was previously an experimental process that took years and hundreds of thousands of dollars per protein. Understanding the 3D structure of proteins will lead to new targeted drugs.[19] Or to take another example, recently it took researchers at IBM and the University of Oxford a matter of weeks to train a generative AI with general information about proteins to identify potential antivirals for COVID-19, synthesize, manufacture, and test against the virus.[20] More broadly, foundational AI stands to rewrite how science is conducted, which includes using LLMs to help predict discoveries in physics or biology, formulating better hypotheses for testing, and conducting faster, cheaper and larger experiments.[21]

The AI opportunity for international trade

AI is also impacting international trade in various ways, and LLMs bolster this trend.[22] Where AI improves worker and firm productivity, this should lead to more trade as firms are more competitive.[23] Indeed, it is already the case that firms

[17] Webb M. (2020). The impact of artificial intelligence on the labor market. Working Paper, Stanford University. Accessed 18 September 2023. Available from URL: https://www.michaelwebb.co/webb_ai.pdf
[18] John Jumpter, et al., "Highly accurate protein structure prediction with AlphFold, Nature, 15 July 2021
[19] Science's 2021 Breakthrough of the Year: AI brings protein structures to all | Science | AAAS
[20] Kenna-Hughers-Castleberry, "AI can suggest Covid-19 antivirals from protein sequence alone." https://www.chemistryworld.com/news/ai-can-suggest-covid-19-antivirals-from-protein-sequence-alone/4017651.article.
[21] Eric Schmidt, "This is How AI will transform the way science gets done," MIT Technology Review, July 5, 2023
[22] Joshua P. Meltzer, The Impact of AI on International Trade. https://www.brookings.edu/articles/the-impact-of-artificial-intelligence-on-international-trade/.
[23] Marc J Melitz and Stephen J Redding, "Heterogenous Firms and Trade" 2014, hand of International Economics, 4th Ed. 1-54 (Elsevier), Martin N Bailey, Eric Brynjolfsson, Anoton Korinek, "Machines of mind: The case for an AI-powered productivity boom", Brookings May 10, 2023 https://www.brookings.edu/articles/machines-of-mind-the-case-for-an-ai-powered-productivity-boom/

most adept at using AI are more productive than non-AI-adopting firms.[24] AI can help firms analyze data to better forecast demand in other countries. AI can also help optimize production and logistics, inform decisions about pricing, inventory levels, and market trends. AI can also aid in identifying new markets for products and services and developing new products and services that are tailored to the needs of specific markets. These capabilities will allow businesses to expand their reach and grow their sales. AI will also be used to optimize the efficiency of global value chains. For example, AI provides the opportunity to increase automation and improve inventory management. Meanwhile, better analysis of overseas demand should allow for more efficient supply chains.

Foundational AI can also reduce trade costs that are a barrier to services trade. For instance, AI-enabled translation services can reduce the costs of trade in services in different languages. As a result of eBay's machine translation service, eBay-based exports to Spanish-speaking Latin America increased by 17.5 percent (value increased by 13.1 percent).[25] PaLM 2—Google's LLM—has multilingual proficiency and translation capabilities in over 100 languages.[26]

AI will also create opportunities to use e-commerce platforms for international trade. For small businesses in particular, digital platforms have provided unprecedented opportunities to go global. In the U.S., for instance, 97 percent of small businesses on eBay export, compared to just 4 percent of offline peers. AI will expand the utility that platforms provide for small businesses to engage in international trade. This will include better analysis of customer data, including browsing history, purchase behavior, and preferences that can make personalized product recommendations.[27]

AI can also enable more efficient and targeted trade finance.[28] AI can analyze vast amounts of data, including financial records, market trends, and customer behavior, to assess creditworthiness, detect fraud, and manage trade-related risks more effectively.

[24] Dirk Czarnitzki, Gaston P. Fernandez and Christian Rammer, Artificial Intelligence and firm-level productivity, J. of Econ. Behavior & Org. Vol 211, July 2023, 188-205
[25] Brynjolfsson, E, X Hui, and Meng Liu (2018), "Does Machine Translation Affect International Trade? Evidence from a Large Digital Platform
[26] Brynjolfsson, E, X Hui, and Meng Liu (2018), "Does Machine Translation Affect International Trade? Evidence from a Large Digital Platform
27 eBay 2015. "Empowering People and Creating Opportunity in the Digital Single Market" An eBay report on Europe's potential, October 2015.
[28] Dharmarajan Sankara Subrahmanian, Artificial Intelligence Platforms Will Drive the Next Phase of Trade Finance Growth, Forbes, Dec 20, 2022, https://www.forbes.com/sites/forbestechcouncil/2022/12/20/artificial-intelligence-platforms-will-drive-the-next-phase-of-trade-finance-growth/?sh=c16cde63b046

Trade facilitation is another area where AI is expected to have a positive impact, complementing efforts to digitize trade documents.[29] AI-powered systems can analyze trade documents, verify product compliance with regulations, detect fraudulent activities, and improve risk-based targeting of commercial shipments.[30] This will help to reduce administrative burdens, enhance security, and lead to better compliance with international trade rules.

Yet geopolitics will lead to reduced trade in AI with China

Developments in AI and foundational AI models are already part of U.S.-China geopolitical competition as both countries race to ensure they lead AI innovation and shape the governance of AI. [31] China is intent on being a global leader in AI, and its 2017 New Generation AI Development Plan lays out steps to 2030 when China will be the world's primary AI innovation center.[32] China is also very capable in AI, by all accounts second only to the U.S.[33] There is also significant foreign investment in Chinese AI startups, as the second largest AI market behind the U.S., between 2015-2021.[34]

This competition over AI has already spilled over into U.S.-China trade and investment flows, driving so-called de-risking of the U.S. (and allied) economies from China in areas of critical technology, including AI. On October 7, 2022 and October 22, 2023, the Biden administration imposed comprehensive restrictions on exports to China of advanced semiconductors needed for AI applications, and the software and equipment needed to make semiconductors.[35] The U.S. has also prohibited engineers and scientists from assisting China in developing advanced semiconductors. In addition, the U.S. has tightened investment screening by Chinese investors into critical technology in the U.S. including AI, and most

[29] White Paper on the use of Artificial Intelligence in Trade Facilitation, UNECE, February 2023 https://unece.org/sites/default/files/2023-02/WhitePaper_AI-TF_Feb2023_0.pdf
[30] WTO/WCO Study Report on Disruptive Technologies, June 2022
[31] Ian Bremmer and Mustafa Suleman, The AI Power Paradox, Foreign Affairs, Sept/Oct 2023, Sisson, M., 2023. Artificial Intelligence, Geopolitics, and the US-China Relationship, Konrad-Adenauer-Stiftung. Germany. Retrieved from https://policycommons.net/artifacts/3375807/artificial-intelligence-geopolitics-and-the-us-china-relationship/4174654/ on 06 Nov 2023. CID: 20.500.12592/3pd5z7.
[32] "Notice of the State Council on Issuing the New Generation Artificial Intelligence Development Plan" [国务院关于印发新一代人工智能发展规划的通知], PRC State Council, 2017, https://perma.cc/B9ZR-5LQL
[33] Kerry, Meltzer, and Sheehan, Can Democracies Cooperate with China on AI Research, Brookings Working Paper, Jan 9, 2023. https://www.brookings.edu/research/can-democracies-cooperate-with-china-on-ai-research/
[34] Emily S. Weinstein and Ngor Luong, "U.S. Outbound Investment into Chinese AI Companies", CSET, February 2023
[35] https://www.bis.doc.gov/index.php/documents/about-bis/newsroom/press-releases/3158-2022-10-07-bis-press-release-advanced-computing-and-semiconductor-manufacturing-controls-final/file; https://www.bis.doc.gov/index.php/documents/about-bis/newsroom/press-releases/3355-2023-10-17-bis-press-release-acs-and-sme-rules-final-js/file

recently issued an executive order to come into effect in 2024 that would prevent U.S. outbound investment into key technology sectors in China, including AI.[36]

The net result is that geopolitical competition with China is reducing international trade and investment between the U.S. and China in the technology and AI compute needed for developing foundational AI models.

[36] https://www.whitehouse.gov/briefing-room/presidential-actions/2023/08/09/executive-order-on-addressing-united-states-investments-in-certain-national-security-technologies-and-products-in-countries-of-concern/

The risks from LLMs

As outlined, foundational AI models such as LLMs present a range of significant economic and trade opportunities.[37] However, to be ambitious and realize these upsides will also require addressing the risk of harm from AI. In other words, ensuring that LLMs are responsible and trustworthy is paramount. This notion of responsible and trustworthy AI picks up on goals expressed in the 2023 Bletchley Declaration on AI Safety agreed by 28 countries and the EU, including the U.S., China, Germany, France, Japan, Indonesia, Brazil, and others, which calls for AI that is "trustworthy and responsible."[38] Effective AI governance that produces responsible and trustworthy AI will be needed to underpin broad-based uptake of AI by governments, businesses, and households.

Many of the risks of LLMs, such as disinformation and risks to privacy, are not new or specific to AI, but may be made more acute. For instance, AI could lead to more misinformation, but this is already a problem with online platforms. LLMs may also introduce new risks, such as risks of exclusion where LLMs are unavailable in some languages. Some of these risks may also end up being mitigated by AI as LLMs are further refined and models become more powerful and accurate. For example, ChatGPT4 is 70 percent more accurate than ChatGPT3.5.[39] That said, ChatGPT4 retains various limitations of associated risks of harm, including bias and misinformation.[40] Moreover, while refining LLMs can reduce some risks of harm, other risks may become more acute as a result. For example, more accurate LLMs can increase the risk of over-reliance by people on the results of LLMs, underscoring that addressing AI risks will involve trade-offs.

A larger point is that developing trustworthy and responsible AI should be in everyone's interest. It is needed as a key building block for optimizing the upsides of AI. However, achieving trustworthy and responsible AI will also require navigating various trade-offs, where optimizing for some value may require sacrifices elsewhere. How this is done and where these trade-offs are struck will require broad-based and inclusive discussions at domestic and international levels. Developing new trade commitments and progress in international economic forums will be an important part of these international efforts. The following outlines the key risks of LLMs to be clear about the challenges before getting into how trade policy and cooperation in international economic forums can realize the upsides and address the risks.

[37] Markus Anderljung and Julian Hazell, "Protecting Society from AI Misuse: When are Restrictions on Capabilities Warranted?",
[38] Bletchley Declaration, https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023
[39] ChatGPT4 Technical Report, 27 March 2023
[40] ChatGPT4 Technical Report, 27 March 2023

Discrimination, exclusion, and toxicity

LLMs are trained on data that encodes existing social norms, with all their biases and discrimination. LLMs will encode unfair discrimination when the data on which it is trained reflects historical patterns of discrimination. For example, earlier versions of ChatGPT4 associated homemaker or nurse with the female pronoun she.[41] When ChatGPT3 was asked to complete a sentence about Muslims, 66 percent of the time it featured Muslims committing violence.[42] Moreover, as LLMs have the capacity for emergent behavior as they scale and learn in the wild, this can lead to different forms of harm over time, and addressing these risks will likely require ongoing assessments of the LLM. However, even here the extent that ChatGPT4 exhibits emergent capacity is uncertain.[43]

LLMs can also risk further marginalization and exclusion of people or groups of people. This can happen when the accuracy of LLMs declines for disadvantaged and marginalized groups that may be using slang or dialects that the LLM does not recognize. As LLMs are more widely used, failing to respond accurately to language prompts can affect access to a wide range of services.

The use of toxic language is a widespread problem with online platforms that may be exacerbated by LLMs. This is also one area, however, where AI can help reduce toxicity, both by identifying and removing it and using technical responses such as human feedback reinforcement. That said, what is toxic language for some is not for others, and context matters, underscoring the challenge. This difficulty of getting toxicity to zero also points to a need to understand what is an acceptable level of risk. Is it zero, is it better than the status quo, or something else? Determining the risk that a country or society is willing to accept is a core expression of sovereignty. However, an explicit discussion about the level of risk tolerance seems necessary.

[41] Bolukbasi, T., Chang, K.-W., Zou, J. Y., Saligrama, V. & Kalai, A. T. in Advances in Neural Information Processing Systems Vol. 29, 4349–4357 (NeurIPS, 2016)
[42] A. Abid, M. Farooqi, J. Zou, "Large language models associate Muslims with violence", Anti-Muslim Bias in GPPT-3, August 2020
[43] Ryan Schaeffer et al, "Are Emergent Abilities of Large Language Models are Mirage", 28 April, 2023 arXiv:2304.15004v1

Security and privacy

Information hazards arise when LLMs disseminate information that is true and can be used to create harm to others. Examples of information hazards include information on how to build a bomb or commit fraud.[44] A related challenge is preventing LLMs from revealing personal information about an individual that risks harming privacy.

Another higher risk from the misuse of LLMs is an increase in the incidence and effectiveness of crime. For instance, criminals can use LLMs to fine-tune spam emails to impersonate an individual, allowing for more targeted manipulation and more successful phishing.[45] This underscores a broader point about the types of risk mitigation techniques that will need to be developed for LLMs, which includes strengthening the human capacity to review and challenge the information provided by LLMs.

Misinformation

LLMs can also be expected to make false statements and reasoning errors, referred to as hallucinations.[46] This remains true for ChatGPT4, though as discussed, with significant improvements over ChatGPT3.5.[47] Given the way that LLMs work by assigning a probability to what should be the next best word based on the previous word, sentence, and overall text, nothing about this presumes the truth of the resulting sentence. In addition, training data drawn from the web contains lots of false statements. Even training LLMs on only factual data would not necessarily overcome this problem as context matters. For instance, a factual statement such as "John owns a car" may be true in one context and not another. LLMs so far do not reliably distinguish between such contexts.[48]

LLMs also increase the risk of greater and more effective misinformation and disinformation campaigns. For instance, LLMs can be used to generate very believable false statements, images, and videos that expand the disinformation space and the harm already caused by online misinformation and disinformation.[49]

[44] N. Bostrom et al, Information Hazards: A typology of potential harms from Knowledge, Review of Contemporary Philosophy, 2011
[45] Markus Anderljung and Julian Hazell, "Protecting Society from AI Misuse: When are Restrictions on Capabilities Warranted?",
[46] G. Branwen, GPT-3 Creative Fiction https://gwern.net/gpt-3
[47] ChatGPT4 Technical Report, 27 March 2023
[48] L. Weidinger et al "Ethical and social risks of harm from Language Models, DeepMind https://arxiv.org/abs/2112.04359
[49] Zellers, R., Holtzman, A., Rashkin, H., Bisk, Y., Farhadi, A., Roesner, F. and Choi, Y., 2019. Defending against neural fake news. Advances in neural information processing systems, 32.

Relatedly, LLMs can also be used by authoritarian governments to improve domestic surveillance and as a propaganda tool.[50]

Overconfidence in the results

There is a related problem with overconfidence in results generated by LLMs. This happens when people anthropomorphize LLMs, overestimate their competencies, and place unwarranted trust in the AI. This is likely to occur as interaction with LLMs appears human-like, passing the Turing test and leading people to assign impressions of warmth and competence (and even consciousness) to AI systems.[51] Overconfidence in the output of such human-like LLMs can lead to even greater reliance on LLMs, including false information, which can perpetuate and expand the scope for harm. Such harm can also be material, such as where it leads people to misdiagnose using LLMs or to base action on information provided by LLMs that is incorrect.[52]

Explainable and interpretable results

LLMs make achieving explainability and interpretability a particular challenge due to the inherently unknowable process of how LLMs produce results and the difficulty measuring the capabilities of these AI models.[53] Explainability requires describing how AI systems function and interpretability is about describing why the LLMs made that particular output.[54] For this reason, it has been noted that foundational LLM can "increase human knowledge but not human understanding." The difficulty of explaining LLM outcomes can exacerbate other potential LLM harms.[55] For instance, interpretability helps users assess whether an LLM is fair, robust, and trustworthy.[56] Being unable to interpret how or why an LLM produced toxic language or discriminatory outcomes can make detecting such failures harder, thereby increasing scope for harm.

Measuring the risk and accountability of LLMs

[50] Markus Anderljung and Julian Hazell, "Protecting Society from AI Misuse: When are Restrictions on Capabilities Warranted?",
[51] McKee, Kevin R., Xuechunzi Bai, and Susan Fiske. 2021. "Humans Perceive Warmth and Competence in Artificial Intelligence." PsyArXiv. February 26. doi:10.31234/osf.io/5ursp.
[52] Bickmore TW, Trinh H, Olafsson S, O'Leary TK, Asadi R, Rickles NM, Cruz R, Patient and Consumer Safety Risks When Using Conversational Assistants for Medical Information: An Observational Study of Siri, Alexa, and Google Assistant J Med Internet Res 2018;20(9): e11510
[53] F. Doshi-Velez and B. Kim, Towards a Rigorous Science of Interpretable Machine Learning arXiv:1702.08608 [stat.ML]
[54] NIST AI RMF (AI RMF 1.0), p16-17
[55] Henry Kissinger, Eric Schmidt, and Daniel Huttenlocher, "ChatGPT Heralds and Intellectual Revolution", WST Opinion, Feb 24, 2023
[56] . Doshi-Velez and B. Kim, Towards a Rigorous Science of Interpretable Machine Learning arXiv:1702.08608 [stat.ML]

LLMs also introduce new challenges when it comes to measuring AI risk. Foundational AI models such as LLMs bifurcate the AI model developer and the entity that then takes the model and develops it for specific applications. This raises new challenges when it comes to ensuring accountability for the LLM across the value chain, which includes how to assess the risk of an LLM when its ultimate use may be unforeseen by the original LLM developer.[57] As the AI value chain lengthens, this raises the issue of how downstream users can assess risk and where to allocate liability for harm. Relatedly, this will also require addressing when access to the foundational model and its underlying data by third parties may be needed.

Copyright infringement

LLMs raise a host of copyright and patent issues.[58] LLMs are trained on the internet which raises the risks of using a lot of copyrighted material. The outputs from ChatGPT4 may also be similar enough to existing copyrighted work such that this output may infringe copyright. Where LLMs produce new creative output or inventions, there is the question as to whether this can receive copyright or patent protection. For instance, is ChatGPT4 a creator, or is the creator the human prompting the chatbot? Finally, LLMs and other forms of foundational AI systems can copy artists, whatever the medium. For example, you can now listen to Drake covering Colbie Caillat or Michael Jackson covering The Weekend, yet these are all generated by AI systems. The question as to whether this output infringes copyright remains unanswered.

The above analysis of risks from foundational AI does not cover all AI risks, including concerns about AI alignment—how to align the goals of AI with humans, particularly when it comes to superhuman AI or artificial general intelligence (AGI), as well as the use of AI for national security purposes and related cybersecurity challenges.[59] These issues are being discussed in other specific forums and trade agreements, and the G7 and TTC may not be well suited to engage with these types of AI risks.

[57] Alex C. Engler and Andrea Renda, "Reconciling the AI Value Chain with the EU's Artificial Intelligence Act", CEPS, September 2022-03
[58] WIPO Conversation on Intellectual Property and Artificial Intelligence, Revised Issues Paper on Intellectual Property and Artificial Protection, WIP/IP/AI/GE/2-/1/Rev. May 21, 2020,
[59] National Security Commission on Artificial Intelligence, https://www.nscai.gov/wp-content/uploads/2021/03/Final_Report_Executive_Summary.pdf

Part 2: International cooperation and a role for trade policy

Why international cooperation on AI is needed

As outlined in the work of the Forum on Cooperation in AI (FCAI), there are a range of reasons that international cooperation on AI is needed.[60] The development of LLMs underscores and makes even more urgent the need for international cooperation in AI. AI will be governed in the first instance domestically, with governments taking different approaches. International AI cooperation has a role in guiding domestic AI governance, improving the outcomes, and building cooperation and interoperability globally among different approaches to AI governance. The following outlines where international cooperation on AI is needed and how foundational AI makes such cooperation even more important.

- International cooperation is needed to update and develop commonly agreed principles for what is responsible and trustworthy AI in the age of foundational AI models.
- International cooperation is needed to address the externalities and extraterritorial impacts of domestic AI regulation that can lead to higher costs for AI innovation and use in other countries, as well as greater AI risk. Foundational AI heightens the need for international cooperation as it accelerates the pace of AI regulation.
- International cooperation is needed to facilitate learning from experience with AI governance.[61] The rapid uptake and use of LLMs and experience with different approaches to regulating AI is generating learning that should be systematically and globally shared.
- International cooperation is needed to expand opportunities for AI R&D and to access the resources needed to use foundational AI systems. Developing AI models, particularly LLMs such as ChatGPT4 is costly and compute-intensive. The result is that only so many companies and governments can run the most advanced LLMs with implications for concentration in capacity. Greater access to foundational AI models consistent with developing responsible and trustworthy AI is needed to ensure that the economic and social benefits are widely shared.

[60] C. Kerry, J.P. Meltzer, A. Renda, A.C. Engler & R. Fanni., "Strengthening International Cooperation on AI", Brookings Report October 2021. https://www.brookings.edu/wp-content/uploads/2021/10/Strengthening-International-Cooperation-AI_Oct21.pdf

[61] Gillian K. Hadfield and Jack Clark, "Regulatory Markets: The Future of AI Governance", April 2023

The role of trade policy in supporting international cooperation on AI

International trade agreements and the discussions underway in international economic forums such as the TTC, G7, and in the OECD, as well as in FCAI, are important for developing international cooperation in AI. Over the past decade, digital issues broadly have become increasingly central to FTAs and DEAs, and figure prominently in international economic discussions.[62] As this section will outline, FTAs and DEAs support domestic AI regulation as well as international cooperation in AI governance. This includes commitments to cross-border data flows, avoiding data localization, agreement not to require access to source code as a condition of market access, agreement to having privacy regulation and developing interoperability mechanisms. In addition, some trade agreements such as the New Zealand-U.K. FTA, digital economy agreements such as the Digital Economy Partnership Agreement (DEPA), and the Australia-Singapore DEA include specific AI commitments. A range of AI issues have also been taken up in various international economic forums, the main ones being the G7, the TTC, and the OECD. Efforts to develop international AI standards in global standards development organizations such as the ISO/IEC are also important areas for developing international cooperation on AI.

This distributed landscape for international cooperation in AI is potentially a feature rather than a bug as it allows for flexible combinations of countries and other stakeholders, and the ability for agenda priorities to adapt quickly in response to developments in AI. Indeed, the explosion of foundational AI models and LLMs, in particular, has underscored the need for international cooperation to be nimble and adaptive. That said, the current landscape for international cooperation on AI has some downsides. This includes the exclusion of some governments and key stakeholders, missed opportunities where progress made in one set of international discussions is not carried over or reflected in others, and duplication of effort.

Currently, the G7 seems the most likely place for effective discussions on AI, though it is not without its limitations. The G7 has a track record on AI, having had AI issues on its agenda since 2016. While the G7 as a seven-country grouping is not globally inclusive, it does include many countries where getting AI governance right will matter most given the preponderance in these countries of AI compute, tech companies, and AI talent. In addition, each year the country hosting the G7 invites a number of other countries to participate, and the European Commission

[62] Joshua P Meltzer, Supporting the Internet as a Platform for International Trade, Brookings Working Paper 69, February 2014 https://www.brookings.edu/wp-content/uploads/2016/06/02-international-trade-version-2_REVISED.pdf

and the OECD also participate in G7 meetings, which further expands the buy-in of G7 outcomes.

The U.S.-EU TTC is another forum where cooperation on AI may be even more rapid and granular than the G7, given the TTC's bilateral nature and technology-focused agenda. Finding ways for the U.S. and EU to cooperate on AI issues will be a key building block for any effective approach to international cooperation on AI governance. Yet, the TTC's bilateral nature will limit its global impact and the government-to-government format of the group will likely limit its relevance. While AI has been discussed in the G20, geopolitical tensions with China and Russia in particular make it unlikely that the G20 can play an effective role in building international cooperation in AI governance in the foreseeable future, and for this reason, is not discussed further here. China does not participate in the G7 or the TTC. While China is a so-called "key partner" in the OECD, it does not engage in OECD work on AI. China is, however, a party to WTO negotiations on e-commerce. The question of how to involve China in AI governance is beyond the scope of this paper but is clearly important.

As a final point, there is an emerging debate about whether these developments in AI governance are enough, with proposals variously calling for new forms of international cooperation and new international organizations.[63] This paper focuses on the narrower question of how to use existing international economic forums and trade agreements to build international cooperation in AI. One reason for this focus on what is actually happening is that many of the AI governance needs identified by some authors are already being developed or could be developed (more or less) using existing international economic forums and through a more robust turn to trade policy. For instance, some proposals call for developing AI standards, yet as outlined here, there is already important work underway in developing international AI standards, such as for risk management frameworks, standards for mutual recognition, and auditing of AI systems. There are also calls by the U.S., the EU, Japan, and others in the G7 and TTC to expand this standards work and to increase the uptake and use of AI standards in domestic regulation. This is also an area where trade policy could contribute to supporting the development and use of international AI standards.[64]

The following outlines key areas in trade agreements, DEAs, and in other international economic forums where there are existing commitments on AI,

[63] European Commission President von der Leyen, State of the European Union speech 2023, https://ec.europa.eu/commission/presscorner/detail/en/speech_23_4426; Lewis Ho et al, International Institution for Advanced AI", arXiv:2307.046992v2, 11 July 2023; Ian Bremmer and Mustafa Suleyman, "The AI Power Paradox, Foreign Affairs, Vol 102, No5. Sept/Oct 2023

drawing on my recent article in the Asia Economic Policy Review.[65] Section three will discuss what more could be done in terms of new trade commitments to support international AI outcomes that maximize the opportunities and help develop AI governance that also addresses the risks from AI.

The WTO

The WTO rules were agreed well before AI was relevant for international trade and even before the impact of the internet and cross-border data flows became an international trade issue. Yet, the WTO remains relevant for AI. The WTO could become even more relevant upon a successful conclusion of the Joint Statement Initiative (JSI) e-commerce negotiations, which could result in a commitment to cross-border data flows, no data localization, and access to source code, which would likely support easier access to better data for AI projects and reduce developer risk when exporting AI models. The following outlines key WTO rules for AI. Specifically, under the GATS, where WTO Members have made a mode 1 services commitment there is also a commitment to allow for the data flows to deliver that service.[66]

The WTO Agreement on Technical Barriers to Trade (TBT) requires WTO members to use international standards as a basis for their domestic regulation and to justify departures from international standards.[67] These commitments could apply to AI standards for products and be a basis for building interoperability across AI regulation. The TBT Agreement also includes commitments to cooperation on mutual recognition and conformity assessment agreements which can help reduce costs of trade where exporters can avoid the costs of multiple conformity assessment processes for AI.[68] These TBT commitments are, however, limited in that they only apply to goods and not services. Yet, AI systems and LLMs will be deployed in many instances as services in the market via APIs and the cloud.

Under the WTO plurilateral Information Technology Agreements (ITA) I and II, WTO members have also agreed to reduce tariffs on a range of technology products, including some used to support AI development, such as goods used to expand internet connectivity and use. The WTO TRIPS agreement includes an agreement on international intellectual property (IP) standards developed in

[65] Joshua P. Meltzer, The Impact of Foundational AI on International Trade, Services and Supply Chains in Asia, Asian Economic Policy Review, November 2023, https://onlinelibrary.wiley.com/doi/10.1111/aepr.12451
[66] Appellate Body Report, United States—Measures Affecting the Cross-Border Supply of Gambling and Betting Services, ¶ 202, WT/DS285/R, (adopted Apr. 25, 2013); Appellate Body Report, China — Measures Affecting Trading Rights and Distribution Services for certain Publications and Audiovisual Entertainment Products, ¶ 151, WTO Doc. WT/DS363/AB/, (adopted Dec. 21, 2009).
[67] TBT Agreement Article 2.4
[68] TBT Agreement Article 5

various IP treaties. Yet, the key issues raised by foundational AI models such as LLMs are not specifically addressed in these international copyright commitments.

While WTO rules remain relevant for AI, the WTO is unlikely to build the international cooperation on AI that is needed. This reality reflects the larger institutional challenges the WTO faces in addressing new trade issues, and similar to, what hobbles the G20, geopolitical competition over AI will prevent a multilateral forum such as the WTO from making significant progress. For these reasons, FTAs, DEAs, and other international economic forums such as the G7, the OECD, and the TTC will need to be the focus of efforts.

Free trade agreements (FTAs) and digital economy agreements (DEAs)

Access to data

There have been significant developments in FTAs and DEAs that are relevant to AI. A recent development in FTAs is the emergence of Digital Trade Chapters. These chapters now include a range of commitments relevant to AI, such as commitments to cross-border data flows and avoiding data localization measures. These commitments matter for AI as they affect access to data for AI using cloud and APIs.

These commitments come with an exceptions provision. The extent of this exception strikes a balance between the commitment to, for instance, the free flow of data and the degree to which governments can impose restrictions on cross-border data flows to meet other regulatory objectives. For example, the CPTPP, USMCA, U.K.-Japan Comprehensive Economic Partnership Agreement and the New Zealand-U.K. FTA include commitments to cross-border data flow and to no data localization measures along with an exception provision modeled on the GATS general exception provision in Article XIV. In contrast, the exception provision in the Regional Cooperative Economic Partnership (RCEP) to the commitment to cross-border data flows is based on GATS Article XIV bis national security exception, allowing for much broader government discretion to restrict data.

Access to source code

Modern trade agreements and DEAs also include a commitment not to require access to source code as a condition of market access. Control over source code is a key source of value and can determine control of the AI model. The CPTPP, USMCA, and the Australia-Singapore DEAs include commitments not to require access to source code as a condition of import. These commitments are also balanced again the need for access by the

government for regulatory purposes. For example, USMCA preserves the rights of regulatory or judicial bodies to require access to source code for a specific investigation, inspection, enforcement action, or judicial proceedings subject to safeguards against unauthorized disclosures.[69]

Interoperability

Another focus of trade policy and international economic cooperation more broadly is on developing interoperability mechanisms. Interoperability is focused on enabling cross-border data flows given different approaches to data regulation. This matters for AI development given the importance of data for AI and LLMs in particular. For example, CPTPP states that the parties will "encourage the development of mechanisms to promote compatibility between these different regimes. These mechanisms may include the recognition of regulatory outcomes, whether accorded autonomously or by mutual arrangement, or broader international frameworks."[70] USMCA states that each party should encourage the development of mechanisms to promote compatibility between these different regimes. The parties to USMCA "recognize that the APEC Cross-Border Privacy Rules system is a valid mechanism to facilitate cross-border information transfers while protecting personal information"—another way of saying that this is an interoperability mechanism.

Open government data

Related to the importance of access to data for AI, trade agreements increasingly include a commitment to open government data. Governments possess considerable amounts of data, whether in the form of tax returns, medical records, or meteorological data. All of this data has potential use cases in training AI systems. The move to make government data more accessible therefore matters for AI. For example, the USMCA digital trade chapter includes provisions on the availability of government data. [71]

AI-specific commitments

The New Zealand-U.K. FTA has notably gone further than other FTAs with respect to making specific AI commitments in the digital trade chapter. This includes an agreement to account for principles and guidelines of relevant international bodies when developing AI governance frameworks to take a risk-based approach to AI regulation that acknowledges industry-led standards development and risk management best practices. Other areas

[69] USMCA Article 19.16
[70] CPTPP Article 14.8
[71] USMCA Article 19.18; CPTPP

of AI cooperation include enforcement, cross-border research and development, and algorithmic transparency.

The Australia-Singapore Digital Economy Agreement and the Digital Economy Partnership Agreement (DEPA) are also starting to directly address AI in the context of ethical use, standards development, talent, and more. The parties to DEPA have agreed to endeavor to promote ethical governance frameworks that support the trusted, safe, and responsible use of AI technologies and to take into consideration internationally recognized principles, including explainability, transparency, fairness, and human-centered values.[72] In the Australia-Singapore DEA, the parties have agreed to share research and industry practice around AI technologies and their governance, to promote the responsible use of AI technologies, and collaborate in the development and adoption of AI governance frameworks that support trusted, safe, and responsible use of AI technologies, taking into account international principles or guidelines on AI governance.[73]

International AI standards

Some FTAs and DEAs have also included a limited commitment to AI standards development and use.[74] The New Zealand-U.K. FTA and Australia-Singapore DEA for instance include commitments to participate in the development of AI standards in regional and international bodies, share experience developing standards, exchange views on potential future areas to develop and adopt standards, and build cooperation with industry on research projects that can increase understanding of the AI standards needed.[75]

[72] DEPA Article 8.2
[73] Australia-Singapore DEA Article 31
[74] Australia-Singapore DEA Article 31
[75] Australia-Singapore DEA Article 30

Other international economic forums

As already touched on, there are a range of other international economic forums where cooperation on AI is being developed. The key ones are the G7, the U.S.-EU Trade and Technology Council, and the OECD. This is not a complete overview of the forums for international discussion on AI, which also include work by the U.N. to develop a Global Digital Compact and in the Global Partnership on AI (GPAI). The Indo-Pacific Economic Forum, GPAI, and the Quad are also involved in different ways with developing cooperation on AI but are not addressed further here as they have yet to lead on AI governance in the way that has been seen with the G7, TTC, and OECD. The G20 is another international economic forum where AI and AI-related issues have been discussed. However as noted earlier, the role of the G20 is not addressed here as geopolitical competition with China over AI and the inclusion of Russia makes G20 progress on AI issues unlikely.

The G7

The G7 is emerging as a key venue for leadership on a range of digital policy issues including AI. Most recently, in 2023 G7 leaders established the Hiroshima AI process with a focus on generative AI.[76] On October 30, 2023 the G7 released International Guiding Principles for Organizations Developing Advanced AI Systems and an International Code of Conduct for Organizations Developing Advanced AI Systems, both documents covering foundational AI models.[77] The Guiding Principles updates the OECD AI Principles to consider new risks posed by foundational AI models. The Code of Conduct builds on the Voluntary AI Commitments large tech companies made at the White House in July and is a set of steps companies agree to take to "seize the benefits and address the risks and challenges brought about by these technologies."

[76] G7 Hiroshima Leaders Commuique.
[77] https://digital-strategy.ec.europa.eu/en/library/hiroshima-process-international-guiding-principles-advanced-ai-system

The G7 has also been central in developing the G20 notion of "data free flow with trust" (DFFT). This includes in 2023 the G7 agreement to establish an Institutional Arrangement for Partnership (IAP) to progress DFFT. Relatedly, the G7 has also developed the importance of interoperability as a means of enabling DFFT. The 2023 G7 Digital and Tech Ministers Statement also reaffirmed the importance of developing interoperability mechanisms, specifically with respect to AI governance frameworks. Under the G7 2023 Digital and Tech Track the G7 will:[78]

- Raise awareness of international AI technical standards.
- Build capacity among stakeholders to participate in the development of international AI technical standards.
- Encourage the adoption of international AI standards as a tool for advancing trustworthy AI.

The G7 has also been leading the development of principles on digital trade that can also support AI. In 2021 the G7 released G7 Digital Trade Principles, which includes the principle that "data should be able to flow freely across borders with trust" and elaborates on how to balance opportunities from data flows with the need for domestic regulation that might restrict cross-border data flows.[79] This includes an agreement to "address unjustified obstacles to cross-border data flows, while continuing to address privacy, data protection, the protection of intellectual property rights, and security."[80] The 2021 G7 Digital Trade Principles also recognize the need to "cooperate to explore commonalities in our regulatory approaches and promote interoperability between G7 members."[81]

The US-EU Trade and Technology Council (TTC)

In the TTC, the U.S. and EU have identified trustworthy and innovative AI as a key priority. Since then, the TTC has made some progress on AI cooperation. The main area is the development of a joint road map with a focus on three areas of cooperation:

- Interoperable definitions of key terms such as trustworthy, risk, harm, risk threshold, and socio-technical characteristics such as bias, robustness, safety, interpretability, and security. A shared and

[78] https://g7digital-tech-2023.go.jp/topics/pdf/pdf_20230430/ministerial_declaration_dtmm.pdf
[79] G7 Digital Trade Principles
[80] G7 Digital Trade Principles
[81] "G7 Trade Ministers' Digital Trade Principles," GOV.UK, October 22, 2021, https://www.gov.uk/government/news/g7-trade-ministers-digital-trade-principles.

consistent understanding of these concepts and terminology is key for operationalizing AI and risk management in an interoperable fashion.

- Support for multi-stakeholder development of AI standards, including cooperation on AI standards development, convening stakeholders to promote representation in SDOs, promoting the development and use of international AI Standards, and developing technical tools to map, measure, manage, and govern AI risks. The U.S. and EU also agreed to adhere to the WTO TBT principles, i.e., to use international standards as appropriate as the basis for technical regulations, conformity assessment, and regional standards.[82]
- Monitor and measure existing and emerging AI risks, including developing a tracker of risks and risk categories that can provide common ground for the U.S. and EU to better define risks and their impact.

The May 2023 TTC Ministerial produced the EU-U.S. Terminology and Taxonomy for Artificial Intelligence, a list of 65 AI terms.[83] This includes technical terms such as "synthetic data" and "reinforcement learning" as well as more socio-technical terms such as what is meant by "accuracy," "human-centric AI," and "resilience." These terms are a “first edition” and open to feedback and further revision. Alignment in AI terms is a necessary building block to more robust cooperation on international standards for trustworthy AI. Developing a shared understanding of these terms is needed as a building block toward developing a common approach to AI standards, regulations, and policies. Getting broader agreement on key terms can help align domestic AI regulation and underpin international cooperation on auditing to support the development of international AI standards.

The U.S.-EU TTC is also engaging in open government data. This includes identifying and promoting best practices for open government data, facilitating collaboration between government agencies, businesses, and civil society organizations on open government data, supporting research and development on open government data, and promoting international standards for open government data.

Standards development organizations

[82] TTC Joint Roadmap on Evaluation and Measurement Tools for Trustworthy AI and Risk Management, December 1, 2022.
https://www.nist.gov/system/files/documents/2022/12/04/Joint_TTC_Roadmap_Dec2022_Final.pdf
[83] https://ec.europa.eu/commission/presscorner/detail/en/statement_23_2992

There is already significant activity underway in various domestic, regional, and global standards development organizations (SDOs) on AI technical and socio-technical standards. AI standards are being developed in SDOs such as the International Organization for Standardization (ISO), the Institute of Electrical and Electronics Engineers (IEEE), and the International Telecommunication Union (ITU). This includes AI standards around concepts and terminology (ISO/IEC 22989) and AI risk management systems (ISO/IEC 42001 and 23894). There is also a range of AI standards under development on data quality management and governance, AI system testing, and oversight of AI systems. For instance, the IEEE has a draft standard for Algorithmic Bias Considerations, a draft standard addressing the record-keeping requirements in the EU AI Act, and a Standards Model Process for Addressing Ethical Concerns During Systems Design.[84]

A defining feature of global SDOs such as the IEC, ISO, and IEEE is that they are multi-stakeholder and industry-led. Governments and civil society participate alongside the private sector. International standards developed by global SDOs are typically based on consensus and are voluntary, in that it remains up to governments and businesses whether to use them. Yet, despite their voluntary nature, international AI standards developed by global SDOs will likely have significant effects on AI. AI developers are likely to use AI standards as benchmarks in contracts and as a basis for industry self-regulation. Governments are also likely to reference AI standards in domestic laws or regulations, making them de facto binding. Indeed, the EU Act will rely extensively on AI standards in areas such as risk management systems, governance and quality of data sets, record keeping, human oversight, and post-market monitoring. Under the EU AI Act, conformity with AI standards will create a presumption of conformity with the Act. The NIST AI RMF also references multiple AI standards from global SDOs.

The importance of international cooperation on standards, as well as the role of international standards in minimizing unnecessary regulatory diversity that can segment markets and raise costs of compliance, has long been a feature of trade policy. As outlined, the WTO TBT agreement, which is also reflected in FTAs includes commitments to base domestic regulation on international standards. When it comes to AI, the development of international AI standards in global SDOs provides an opportunity to use trade policy to reinforce the importance of cooperation on AI standards and to using international AI standards as a basis for domestic regulation.

[84] IEEE P7003, IEEE P7001, IEEE 7000

Part 3: Next steps for trade policy and discussions in international economic forums

As discussed, foundational AI models heighten the need for international cooperation on AI, particularly in light of the speed at which LLMs like ChatGPT4 are being developed and adopted. Indeed, the call for a moratorium on further versions of ChatGPT4 applications and research speaks to growing anxiety.[85] As outlined, there are already important commitments on digital trade that matter for AI, and AI is a focus of discussion in a range of international economic forums. The rapid pace of AI development, the learning needed to understand the opportunities and risks of AI, as well as the need to develop best practices when it comes to AI regulation require a strategic two-tiered, mutually reinforcing role for trade agreements and discussion in international economic forums. Trade agreements should elevate the output from AI-focused forums and standards bodies into trade commitments and develop new commitments. International economic forums such as the G7, the TTC, and the OECD also provide opportunities for sharing regulatory experience and testing new forms of cooperation on AI that could be later ripe for inclusion in trade agreements. FCAI as a track 1.5 dialogue is another forum to explore cutting-edge AI issues. The following outlines where additional commitments in trade agreements and DEAs are also needed and where to build on the AI-focused discussions in the various international economic forums.

Access to AI compute

AI compute covers the hardware and software that supports AI workloads and applications.[86] Access to AI compute is critical if countries are to develop foundational AI and LLMs. Yet the AI compute needed to run foundational AI models keeps growing rapidly. By some estimates, the computational capacity required to train AI models has grown by hundreds of thousands of times since 2012.[87] For instance, training ChatGPT4 has required access to supercomputers using state-of-the-art hardware (CPUs and GPUs) and high-bandwidth networks that access top cloud infrastructure.[88] AI platforms or software are also needed to build or implement AI capabilities, such as TensorFlow or PyTorch, as well as the applications to deliver AI capabilities.

[85] Pause Giant AI Experiments: An Open Letter, March 22, 2023 https://futureoflife.org/open-letter/pause-giant-ai-experiments/
[86] OECD.AI Expert Group on AI Compute and Climate
[87] Sevilla, J. et al. (2022), "Compute Trends Across Three Eras of Machine Learning", https://arxiv.org/abs/2202.05924
[88] Microsoft announced new supercomputer, lays out vision for future AI, May 19, 2020, Microsoft announces new supercomputer, lays out vision for future AI work - Source

According to one survey of 450 industry professionals in the U.S. and Europe, access to AI compute is now the key challenge facing AI development, surpassing access to data.[89] In the U.S., the NAIRR Task Force highlighted the extent that the AI R&D ecosystem in the U.S. is becoming inaccessible for many businesses and researchers. These developments—the growing cost of AI compute needed to train LLMs—point to the need for expanding AI capacity.[90]

Trade policy can support the development of access to AI compute and data by reducing barriers to AI infrastructure, data, and cloud computing as well as AI services. In some cases, this may be about reducing trade barriers to the hardware needed for AI compute. In other cases, it is about reducing barriers to trade in services that are needed to access AI compute and AI services themselves. For instance, Turkey's prohibition on use of cloud computing services by public institutions and Korea's cloud security requirements create barriers to trade in cloud services that can negatively affect the development and uptake of AI.[91] Commitments in trade agreements on avoiding data localization measures could get at some of these barriers and highlight their relevance for AI.

Risk-based AI regulation

One area where trade policy could be developed further is by giving added content to existing international agreement that AI regulation will be risk-based. As noted, the New Zealand-U.K. FTA includes a commitment to a risk-based approach to AI. The 2023 TTC ministerial affirmed the importance of a risk-based approach.[92] In addition to the EU and the U.S., various other governments are developing risk-based approaches in their AI regulation, including Japan, the U.K., Canada, and Brazil. More is needed on what it will mean for regulation to be risk-based, including what are the risk assessment and risk management tools that governments develop, and organizations adopt. The NIST AI RMF is one example of how organizations can go about conducting a risk assessment for AI that could be used globally.[93] The AI RMF also references international AI standards, making the AI RMF a strong candidate for building interoperability among AI regulations calling for a risk-based approach to AI. Trade agreements could incorporate or

[89] Run: AI's 2023 State of AI Infrastructure survey reveals that infrastructure and compute have surpassed data scarcity as the top barrier to AI development (prnewswire.com). https://www.prnewswire.com/news-releases/runais-2023-state-of-ai-infrastructure-survey-reveals-that-infrastructure-and-compute-have-surpassed-data-scarcity-as-the-top-barrier-to-ai-development-301746292.html
[90] A Blueprint for Building National Compute Capacity for Artificial Intelligence, OECD Digital Economy Papers, February 2023, No. 350
[91] 2023 NTE Report.pdf (ustr.gov). https://ustr.gov/sites/default/files/2023-03/2023%20NTE%20Report.pdf.
[92] U.S.-EU Joint Statement of the Trade and Technology Council | The White House
[93] https://www.nist.gov/itl/ai-risk-management-framework

reference the AI RMF as an agreed tool. The TTC and G7 could also reference the AI RMF as an example of an approach to a risk-based approach to AI regulation. [94]

There are other ways that FTAs and DEAs can develop commitments to risk-based AI regulation. The WTO Sanitary and Phytosanitary (SPS) Agreement provides some guidance here. A key commitment in the SPS Agreement is that governments undertake risk assessments and base their SPS measures on risk assessments.[95] Other relevant SPS commitments are that regulations are not more trade restrictive than necessary to achieve the appropriate level of SPS protection.[96] Under the SPS agreement, governments also remain free to set their own level of risk tolerance. Using the SPS Agreement as a guide, FTAs and DEA could include commitments to base AI regulation on a risk assessment, to specify the risks against which potential harm is to be assessed, and to provide explanations for risk management practices and approaches that in effect regulate AI in ways that are more restrictive than necessary to achieve each government's chosen level of risk tolerance.

Government procurement and responsible and trustworthy AI

Trade agreements could include commitments on government procurement that supports the development of responsible and trustworthy AI. In many countries, government procurement will be an important way to influence how AI is developed. For instance, U.S. government agencies are required to develop regulatory plans for AI, and a number have done so.[97] The U.S. Executive Order on AI directs federal government agencies to develop standards and guidelines and reports to address risks from AI as well as to encourage the uptake and use by the federal government of AI.[98] EU agencies will also need to develop AI policies under the EU AI Act as they assume responsibility for regulating AI incorporated into regulated products. Governments can also seek to drive responsible and trustworthy AI by setting standards through government procurement. Trade agreements can help here by including commitments that government procurement contracts are based on international AI standards and are nondiscriminatory.[99] Commitments such as these would support the uptake and globalization of international AI standards and promote regulatory compatibility

[94] U.S.-EU Joint Statement of the Trade and Technology Council | The White House https://www.whitehouse.gov/briefing-room/statements-releases/2022/12/05/u-s-eu-joint-statement-of-the-trade-and-technology-council/.
[95] SPS Agreement Article 5.1
[96] SPS Agreement Article 5.5
[97] Maintaining American Leadership in Artificial Intelligence, February 2019 EO 13859 and OMB guidance M-21-06
[98] US Executive Order on Safe, Secure and Trustworthy Artificial Intelligence https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/
[99] See for example the NZ-UK FTA, Article 16.4 & 16.9

among countries. The G7 Code of Conduct for AI could also be expanded via its uptake in government procurement contracts. Commitments to nondiscrimination would also support international trade in AI products and services.

Conformity assessment and auditing of LLMs

Assessing compliance of AI systems with AI regulations and standards will require ex-ante conformity assessment mechanisms and ex-post monitoring, as well as auditing of AI systems in goods and services. When AI is exported in products such as medical devices or motor vehicles, mutual recognition agreements (MRAs) between countries of conformity assessment can allow for testing AI products with the importing country's AI regulation in the country of export, reducing the uncertainty and costs of trade. A complementary step is recognition by the importing country of conformity assessment bodies in the exporting country able to undertake the conformity assessment.

There are various efforts underway to develop conformity assessment and auditing systems for AI. The EU AI Act requires ex-ante conformity assessments for high-risk AI systems by third parties referred to in the AI Act as a "notified body." Avoiding such requirements for conformity assessment becoming a trade barrier will require the development of mutual recognition agreements with other countries. The AI Act does seem to foresee MRAs with third countries. Entities responsible for high-risk AI systems must also meet auditing documentation requirements. In the U.S., regulatory authorities such as the Federal Trade Commission are focusing on ex-post oversight of industry self-assessment of compliance with their AI policies.

Various DEAs have made initial progress on building cooperation on MRAs that could apply to AI. For example, the Australia-Singapore DEA includes a recognition of the importance of conformity assessment to support digital trade and includes an agreement to "endeavor to exchange information to facility conformity assessment to support digital trade."[100] Building on this could include new commitments to developing the necessary MRAs and recognition of conformity assessment bodies with respect to AI.

Another related area where trade policy could do more is with respect to auditing foundational AI models. The AI Act requires that third-party conformity assessment bodies carry out periodic audits to ensure that the AI provider has the internal quality management system and to provide an audit report.[101] The proposed amendments by the EU Parliament to the AI Act note the need to develop auditing capacity and call for internal auditing of foundational AI models

[100] Australia-Singapore DEA Article 30.5
[101] Ai Act Annex VII. Clause 5.3

to be broadly applicable, i.e., a common approach to assessing risk across AI systems.[102] It is likely that the auditing of AI systems will become a feature in how the government regulates AI. Auditing may also be needed where AI regulation relies on self-assessment by companies for compliance with laws and regulations and with their own internal standards and processes for delivering trustworthy AI.

To effectively audit foundational AI models may require a tiered and multilayered approach. This could include governance audits that assess the organization developing the AI model, its organizational procedures, accountability structures, and quality management systems. Process audits of the AI model and its datasets, as well as ex-post downstream application audits, may also be necessary.[103]

Enabling effective audits and avoiding audit requirements becoming trade barriers will require common auditing standards and recognition of audit reports carried out in third countries. This can be facilitated by MRAs with third countries that recognize who can qualify as an auditor and what is an audit report for domestic AI regulation. Trade agreements could include commitments to MRAs for auditing. In addition, trade agreements could be used to support domestic uptake of conformity assessment and auditing processes based on international standards. For instance, the ISO/IEC is working on a standard on how to carry out a conformity assessment for AI management systems and the needed competencies for AI auditors. Basing conformity assessment and auditing systems on international standards could enable interpretability of auditing reports across countries, facilitating compliance with domestic AI regulation and building trust in AI systems.

Cooperation on international AI standards

As outlined, there are already some international principles that can guide AI developers, and considerable work is underway in developing AI standards in global SDOs. There are two areas where trade policy can support the development and use of international AI standards. First is by developing new TBT-like commitments that apply to international standards for services, which would cover AI. Second is by supporting the development of international AI standards in global SDOs.

Working to align regional approaches to AI standards with international AI standards

Trade policy in FTAs and DEAs can build on WTO TBT commitments and include a commitment to base domestic AI regulation on international AI standards while providing flexibility to adapt international AI standards

[102] AI Act draft compromise amendments, p. 29 clause (60h)
[103] J. Mokander, et al, "Auditing Large Language Models: A Three-Layered Approach", 16 Feb 2023

where necessary to respond to local needs and conditions. This would require going beyond the TBT agreement and extending the commitment to AI as a service. AI regulation based on international AI standards should also benefit from a presumption of consistency with the trade agreement.

There are, however, limitations with the TBT-style approach in the AI context and in particular, the scope of flexibility the TBT Agreement provides to ignore international standards in favor of domestic/regional standards where the government decides that the international AI standard is not fit for purpose. This is due to the socio-technical nature of many AI standards that seek to address technical AI issues as well as many of the broader societal and rights-based impacts of AI. This means that many of the AI standards being developed under the AI Act, for example, will need to address the risks of AI to EU fundamental rights. For instance, the EU AI Act requires standards to establish a risk management system for high-risk AI systems. There are already global standards dealing with risk management, specifically, ISO/IEC 31000 contains general guidelines on risk management, and the AI-specific ISO/IEC 23894 addresses how organizations manage risk. On the one hand, there is an opportunity here to align the EU approach to risk management under the AI Act with global AI standards. Yet, the ISO/IEC standards which address whether AI systems operate consistently with an organization's standards may not meet the EU's need for a risk management system for the impact of AI systems on European fundamental rights.[104] This raises the prospect that the EU standards bodies conclude that international AI standards are not fit for purpose and require instead a regional approach.

To further strengthen a requirement to base domestic regulation on international AI standards, trade agreements should also include commitments that governments will ensure a domestic standards process that is transparent and open to broad participation, opportunities for all stakeholders to submit comments, and obligations on regulators to provide reasons for their decision. Such an outcome would give confidence that departures from international AI standards were driven by legitimate domestic needs rather than protectionism.

Ensure development of AI standards by SDOs that are fit for purpose

[104] Soler Garrido, J., Fano Yela, D., Panigutti, C., Junklewitz, H., Hamon, R., Evas, T., André, A. and Scalzo, S, Analysis of the preliminary AI standardisation work plan in support of the AI Act, Publications Office of the European Union, Luxembourg, 2023, doi:10.2760/5847, JRC132833.

Commitments to base domestic AI regulation on international AI standards also require agreement on what the standards bodies are that can produce the relevant international standards. The TBT Agreement provides guidance here in Annex 1 which defines standards as being based on consensus and as being developed by a body whose membership is open to the relevant bodies of at least all WTO members. The WTO TBT Committee Principles for the Development of International Standards adds further detail and lists the principles and procedures that should be observed when developing international standards.[105] This includes for instance transparency and openness to all WTO Members. These TBT principles seem relevant today for identifying the standards bodies developing international AI standards.

As outlined, discussions in international economic forums on international standards now also include discussion of the operation of the SDOs themselves, such as expanding participation in SDOs by developing-country governments, industry, and civil society.[106] The TTC is also bringing together the U.S. and EU standards bodies and related organizations to work on metrics and methodologies for measuring AI trustworthiness including risk management methods. However, this is one area where the limited membership of the TTC could restrict the impact of its work, which should aim for global uptake. This suggests at least seeding similar efforts in the G7.

Data governance for AI

There are currently only limited commitments in trade agreements on some of the data governance issues specific to AI that address how better data governance can help minimize risks that the data used to train the AI models can cause discrimination, lead to unfair outcomes, misinformation, and privacy violations. As outlined, DEPA and the Australia-Singapore DEA include commitments to sharing information and cooperation on AI governance Frameworks and to the G7 work on Data Fee Flow with Trust (DFFT), and the Global Cross-Border Privacy Rules (CBPR) Forum seeks to facilitate trusted access to personal data. Trade agreements and DEAs also increasingly include a commitment to protecting the privacy of personal data.[107] This is a good beginning, but more focused cooperation on data for AI and foundational AI is needed. For example, the

[105] Decision of the TBT Committee on Principles for the Development of International Standards, Guides and Recommendations with Relation to Articles 2, 5 and Annex 3 of the Agreement. WTO | Principles for the Development of International Standards, Guides and Recommendations: https://www.wto.org/english/tratop_e/tbt_e/principles_standards_tbt_e.htm
[106] Joshua P. Meltzer, "A Critical Technology Standards Metric, assessing the development of critical technology standards in the Asia-Pacific", Brookings Report, September 2022 https://www.brookings.edu/wp-content/uploads/2022/09/CTSM-Report-Sep-2022_Final.pdf
[107] CPTPP Article 14.8, USMCA Article 19.8, DEPA Article 4.2.

heightened risk from LLMs of discrimination and toxicity from foundational AI also underscores the importance of best practices when it comes to data curation and data governance. This is a complex area, but initial steps could aim to share best practices in terms of how LLMs document their data governance practices, how to incentivize appropriate data governance, and methods and experience with opening data and algorithms to scrutiny. Looking ahead, a better understanding of data needs for foundational AI, including the opportunities for synthetic data, would benefit from international cooperation.

Data governance for AI is also being taken up in international standards bodies, it is part of the NIST AI RMF and there are data governance requirements in the EU AI Act. More robust commitments in trade agreements and DEAs on how to use and reflect international AI standards as they apply to data governance is another way to level up a more consistent and robust approach to data governance for LLMs.

Another area where trade policy could add weight is government access to personal data held by private entities for law enforcement and national security purposes. The question of U.S. government access to such data was at the heart of the Schrems II case that has led the Court of Justice of the European Union to invalidate Privacy Shield.[108] In December 2022, the OECD adopted a set of principles governing government access to personal data.[109] This declaration marks an important development in getting at how to enable data free flow with trust. The declaration's principles balance government access to personal data for the legitimate needs of law enforcement and national security, with the need to also protect privacy consistent with broader democratic norms. One focus for the recently agreed G7 IAP will be to increase awareness of this OECD declaration. This declaration could also be specifically referenced in trade agreements as the basis for a shared understanding of the terms on which governments can access data for national security and law enforcement purposes.

Transparency of reporting on foundational AI use

Building trust in how risks from foundational AI models are being addressed will require that the companies responsible for developing and testing foundational AI be transparent about the steps they take to test and mitigate these risks. In the U.S., large technology companies that are at the forefront of developing foundational models made Voluntary AI Commitments at the White House in July where they agreed to publicly disclose red-teaming and safety procedures in transparency reports and to share information among companies and with

[108] CJEU, Data Protection Commissioner v Facebook Ireal and Maximillian Schrems, C-311/18, 16 July 2020
[109] OECD Declaration on Government Access to Personal Data Held by Private Sector Entities, 14/12/2022 https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0487

governments on advances in frontier capabilities and emerging risks and threats.[110] These Voluntary AI Commitments were subsequently used by the G7 as the basis for the International Code of Conduct for Organizations.

Modern trade agreements include comprehensive commitments by governments to regulatory transparency and due process when it comes to developing regulation affecting international trade, including opportunities for comment and commitments for written responses to comments. Trade agreements and DEAs could build on this approach and reference the G7 Code of Conduct as steps that all developers of foundational AI models would agree to take, whether government or private sector actors.

The G7 Code of Conduct also targets actions that organizations should take to enhance information sharing and disclosure of AI governance and risk management policies. These could be turned from voluntary commitments into binding commitments by way of trade agreements, further strengthening trust in foundational AI models. Moreover, in order for reporting and disclosure to be meaningful across countries will require some agreement on what information should be reported and disclosed and in what form. This is another area where FTAs and DEAs could elaborate.

[110] Voluntary AI Commitments, https://www.whitehouse.gov/wp-content/uploads/2023/09/Voluntary-AI-Commitments-September-2023.pdf

Part 4: Conclusion

AI will have significant implications for how economies grow, what jobs are done, how societies work, and how governments function. The release by OpenAI of ChatGPT4 has highlighted the rapid progress being made in foundational AI, with potentially significant new opportunities for economic growth and human flourishing, but also with new risks. Governments are looking to regulate AI, and this is where much of what matters for AI governance will play out. International cooperation is needed to ensure that the AI governance that emerges is effective, enhances economic and social flourishing, and addresses the spillover and extra-territorial impact of domestic AI regulation.

This paper outlines a role for building international cooperation through trade agreements as well as through the various international discussions on AI happening in the G7, the U.S.-EU TTC, and the OECD. In fact, as this paper outlined there is a lot happening and progress is being made. Yet, there are several areas where international cooperation needs to be deepened and expanded in light of foundational AI. This includes how to align approaches to risk assessment for AI, cooperation on conformity assessment and auditing of AI systems, developing international AI standards, and more.

What seems clear is that the key challenge will be to maximize opportunities to use AI globally while ensuring that AI is responsible and trustworthy. This will require regulating AI to minimizes the risks and build trust in the technology, without stifling AI innovation and access. This is a big governance challenge that governments, industry, and civil society are only beginning to understand how to navigate. Undoubtedly, innovative approaches to domestic regulation and international cooperation will be required. Developing new commitments in trade agreements and DEAs, while also expanding and deepening discussions in international economic forums, present key opportunities for developing flexible and new approaches to international cooperation on AI governance that will be required.

BROOKINGS

1775 Massachusetts Ave NW,
Washington, DC 20036
(202) 797-6000
www.brookings.edu