

A nighttime photograph of the Washington Monument and the U.S. Capitol building. The Washington Monument is a tall, white, obelisk-shaped structure that dominates the center of the image. To its right, the U.S. Capitol building is visible, with its iconic dome illuminated. In the foreground, the Lincoln Memorial is partially visible on the left side, with its columns lit up. The sky is dark, and the overall scene is lit by the warm lights of the buildings.

Making Sense of the Chaos

Practicable Policy Solutions for
Unprecedented Times

GEORGETOWN

Spring 2026

PUBLIC POLICY REVIEW

Volume 31

Table of Contents

The Price of Innovation for Hemophilia	4
<i>Brian DuVal, Stacie B. Dusetzina, and Michelle Chi</i>	
Commodity Market Stabilization Strategies for Latin America and the Caribbean	14
<i>Stefan Doyen</i>	
Designing Effective Social Media Regulations.....	26
<i>Grace Falgoust</i>	
Reciprocal Soft Power Between China and Its Developing Partners	36
An Examination of South-South Cooperation Rhetoric	
<i>Odessa Ikels</i>	
Talking With the Mic Off	48
The Passage of Custodial Interrogation Recording Laws	
<i>William Lowthert and Christopher J. Normile</i>	
Climate Change Adaptation Strategies and Care Ethics in Chicago, Illinois	60
<i>Megan Morrell</i>	
Protecting the Digital Town Square.....	70
A Privacy-Preserving Framework To Combat AI-Powered Social Media Bots	
<i>Florin Schoenborn</i>	
Addressing Insulin Affordability	78
Policy Pathways to Expand Access, Reduce Costs, and Improve Health	
<i>Rachel Williams</i>	
The European Union’s Response to Asymmetric Trade Crises	86
Distributional Conflict, Institutional Constraints, and the Politics of Unity	
<i>Matthew Benjamin</i>	
The Future of Work: The Case for a Robot Tax	100
<i>Nikola D. Misetic</i>	
The Virtuous Insider	108
Why Prediction Markets Need the Traders We Usually Ban	
<i>Amr Yakout</i>	
The Nearshoring Prescription	114
A North American Pharmaceutical Chapter	
<i>Lynzee Lee</i>	
Interpreting Before We Deploy	126
A Precautionary Policy Case for AI Interpretability in the United States	
<i>Bhumika Nehbani</i>	
Women’s Participation as a Determinant of Sustainable Peace	136
A Comparative Analysis of Liberia, Colombia, and Northern Ireland	
<i>Katherine Pitcher</i>	

SENIOR EDITORS

Caitlin Rowley Gallamore
Demitra Tomasides

JUNIOR EDITORS

Alisha Bi
Abby Child
Maxwell Keiles
Nicholas Riccio
Dalia Rodriguez Garnica
Cade Roy
Johnny Willing

LAYOUT ARTIST

Tejas Kaur

DATA VISUALIZATION

Helen Wang
Senior Editor
Jiani Zhao
Senior Editor
Joanna Cho
John Deng
Darren Deng
Franklin Estein
Jun Nishida
Emily Zhao

GPPR LEADERSHIP

David Stout
Editor-in-Chief
John McCabe
Managing Editor
Garreth Hui
Finance Officer

A Letter from the Editors

The Georgetown Public Policy Review is not just a journal. It is a gameplan.

Right now, our countries, our cities, and our governments are facing high stakes challenges: How do we respond to emerging technologies, such as artificial intelligence and prediction markets? If climate change is here, how do we mitigate the harm? Is it possible for countries to minimize economic shocks amidst shifting global trade policy? How should we navigate these questions in the face of war and international conflict?

The urgency is only escalating, and policymakers around the world are working to build and implement solutions that make sense of the chaos.

At the Georgetown Public Policy Review, we are doing the same.

In these pages, you will find actionable policy solutions from the new generation of public policy leaders. Whether analyzing market stabilization strategies in Latin America and the Caribbean, designing US social media regulations, or increasing the efficacy of international conflict negotiation... The authors of this volume are developing evidence-based solutions to the world's most intractable problems.

In the previous edition, the editors explored the impact of the post-Covid landscape, positing "this may be the new normal – at least until the next one." In the 31st volume of the Review, we think critically about this "new normal" and present actionable public policy solutions in the hope of a better tomorrow. We share outside-of-the-box ideas from new voices. Most importantly, we believe that thoughtful policy analysis still has the capability to move the needle towards positive change.

So, thank you for reading the 31st edition of the Georgetown Public Policy Review. In these pages, we think, we strategize, and we create a gameplan for navigating this chaotic world.

We hope you will, too.

Caitlin Rowley Gallamore (MPP '26)
Demitra Tomasides (MPP '26)

Senior Editors, 2026 Spring Edition

Georgetown Public Policy Review

The Price of Innovation for Hemophilia

Brian DuVal¹, Stacie B. Dusetzina², and Michelle Chi³

Hemophilia is a rare inherited bleeding disorder characterized by an inability for one's blood to clot, which causes internal bleeding, joint damage, and an increased risk of injury and mortality. Individuals are diagnosed with hemophilia when they lack any one of 13 blood factors, which are proteins that assist in blood clotting ("Blood Clotting Factor" n.d.). The most common forms of hemophilia are hemophilia A (factor VIII deficiency) and hemophilia B (factor IX deficiency). Symptoms manifest similarly in hemophilia A & B, but treatments differ slightly as different factors need to be replaced (Centers for Disease Control and Prevention n.d.). Hemophilia ranges in its severity. Individuals with mild hemophilia produce 5-40% of the needed factor, individuals with moderate hemophilia produce 1-5% of the needed factor, and individuals with severe hemophilia produce less than 1% of the needed factor (National Bleeding Disorders Foundation n.d.). These differences in factor levels are relevant to an individual's treatment and lifestyle. For example, someone living with mild or moderate hemophilia may often take medication on an as-needed basis, while an individual with severe hemophilia must take medication on a preventative basis to routinely replace what is missing.

Approximately 33,000 people in the US are affected by hemophilia, and it is among the costliest conditions to treat. Average treatment costs are estimated to be \$300,000 per year across all types and severities of hemophilia. However, for individuals living with severe hemophilia who require prophylaxis to prevent spontaneous bleeds, medication costs can reach \$1,000,000/year (Miller 2026). People who live with hemophilia require specialty drugs to replace what is missing from their blood to achieve hemostasis and live a normal life. Hemophilia has had substantial treatment innovation over time, and especially over the past decade. Today, there are over 16 treatments

indicated for hemophilia A prophylaxis and 9 treatments for hemophilia B prophylaxis (National Bleeding Disorders Foundation n.d.). In recent history, the most common treatments for hemophilia involved standard-half-life clotting factors. New treatment options, like extended half-life products and other non-factor options, have transformed the experiences of people with hemophilia (Contreras 2025).

In recent years, products like Genentech's Hemlibra, a factor mimetic product used to treat hemophilia A, and an array of extended half-life (EHL) factor products have given patients more options for treatment. Additionally, new "rebalancing agents," like Pfizer's Hympavzi and Novo Nordisk's Alhemo, provide a new pathway for hemophilia treatment that aims to lower anti-coagulant proteins in individuals with hemophilia A & B, rather than replacing factor (Van Thillo and Hermans 2025). Generally, these newer products may reduce the burden of treatment for patients with hemophilia by requiring less frequent infusions than traditional, standard half-life products. Additionally, they can reduce hospitalizations and emergency department visits and improve the quality of life for patients relative to older therapies due to improved adherence (Contreras 2025).

As of 2023, the US market for hemophilia treatment was valued at \$4.2 billion. As treatments improve and patients live longer, it is expected to grow, reaching \$6.8 billion by 2031 (CMI 2024). The high and recurring costs of treatment are a challenge for insurers, including state Medicaid programs that cover approximately one third of US patients with hemophilia (Kaiser Health News 2018). In this policy essay, we highlight some unique challenges for reducing hemophilia treatment spending and how the current market has provided additional treatment options for

1 Brian DuVal is an MPH student at Vanderbilt University

2 Stacie B. Dusetzina, Ph.D. is a Professor in the Department of Health Policy at Vanderbilt University School of Medicine.

3 Michelle Chi, M.D. is an Assistant Professor of Clinical Medicine, Division of Hematology Oncology at Vanderbilt University Medical Center.

Half-Life Hours of Prophylactic Hemophilia Factor Replacement Products by FDA Approval Year

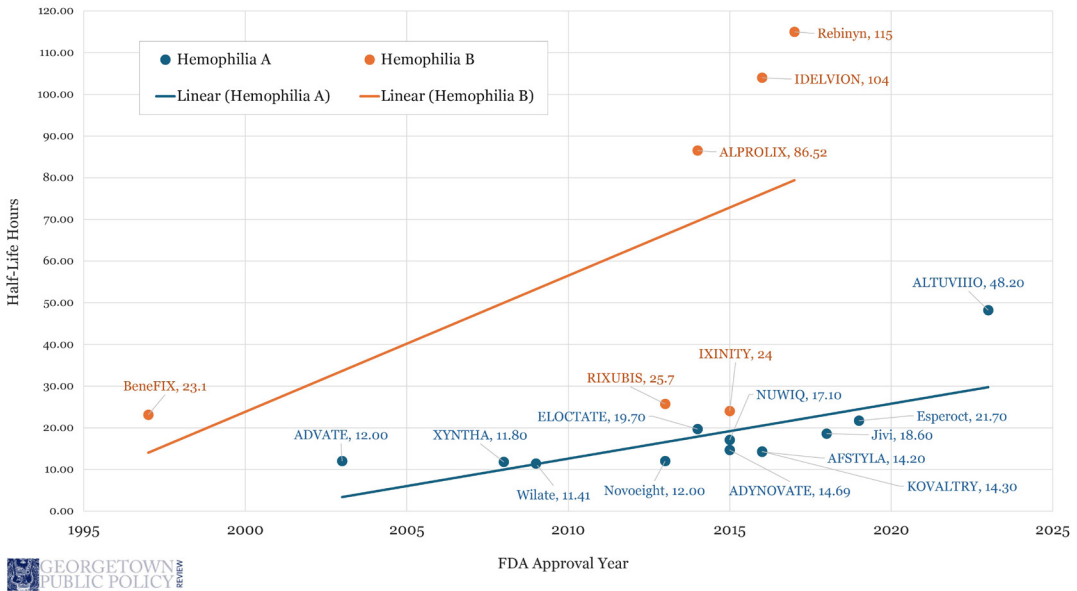


Figure 1: Half-Life Hours of Prophylactic Hemophilia Factor Replacement Products by FDA Approval Year

*Not pictured, HEMLIBRA (645 hour half-life)

patients without increasing price competition among entrants.

The Value of Zero Bleeds

Without treatment, people with hemophilia face substantial morbidity and mortality and have a life expectancy of 11-13 years (Healthline 2023). As the standard of care has improved, patients with access to treatment have been able to live longer and healthier lives, largely free from bleeds and other significant complications associated with hemophilia. Organizations such as World Federation of Hemophilia and National Bleeding Disorders Foundation Medical and Scientific Advisory Council have defined the goal of zero bleeds as the next milestone for the treatment of hemophilia (National Bleeding Disorders Foundation n.d.). A life free from bleeds prevents the worst physical consequences of hemophilia and provides a lifestyle where normal physical ability is achievable.

For some patients, treatments can make this goal a reality. Preventing bleeds lowers hospitalizations, prevents pain and irreversible joint and organ damage, and reduces the

need for additional health care resources. The cost savings for newer treatments—particularly reductions in the frequency of infusions and decreased need for treatment for bleeding events—have largely been priced into hemophilia treatments. Treatments keep getting better, but they keep getting more expensive (Figure 2). This means improved clinical benefits for patients but sustained high spending over time for the treatment of hemophilia.

Innovation and Pricing for Hemophilia Therapies

Despite the number of competitor products for both hemophilia A and B, there is little evidence that drug manufacturers compete on price (Figure 2). Using Medicare’s average sales price (ASP) data in 2024 which considers net prices paid per unit of a drug (factoring in rebates paid by manufacturers), we calculated the expected price per one year of treatment for factor products and non-factor products approved prior to 2024. Estimated costs for prophylaxis were calculated for each product by multiplying the per-unit Medicare ASP by the manufacturer-recommended dose per week for



Figure 2: Estimated Cost of Products Indicated for Prophylactic Treatment of Hemophilia (2024 Medicare ASP)

*ASP = Average Sales Price

*Most factor products have a manufacturer recommended dose of 50 IU/kg. Xyntha (30 IU/kg) Ixinity (70 IU/kg) are outliers on this graph.

*Hemlibra annual cost includes initial loading dose when treatment is initiated

a standard adult patient and multiplied by 52 weeks. Then, we included the cost of a single, on-demand infusion (to represent treatment for a bleed) multiplied by the median overall annual bleed rate from each manufacturer’s clinical trial results and added this to the cost of prophylaxis over the course of a year to arrive at an adjusted, estimated annual cost for each product that takes clinical benefits into account. We standardized dosing to an adult male patient with severe hemophilia with a body weight of 95kg. Roughly 60% of hemophilia patients have the severe form, and this standardized dose represents an adult patient (National Bleeding Disorders Foundation n.d.).

Across all treatments for hemophilia A, average annual prices were \$959,663 (ranging from \$667,372 for Xyntha to \$1,159,418 for Altuviio). Across all treatments for hemophilia B, average annual prices were \$1,015,172 (ranging from \$910,857 for Alprolix to \$1,269,389 for Ixinity). Though prices are slightly higher for newer drugs, differences in prices across drugs were relatively small.

The small patient population and the robust pipeline of new products results in reduced incentives for companies to introduce

lower-priced treatment options (e.g., biosimilar or generic products) even for products that have reached the end of their original patent protection. This allows all products to remain at high prices. Those high prices become benchmarks for a standard of care in the US market and reinforce high prices of future drugs when they are approved.

This can further be illustrated by the growth in costs of most standard half-life factors after the introduction of extended half-life factors between 2014 and 2024 (Figure 3).

The first extended half-life factors, Alprolix & Eloctate, were approved in the US in 2014. These products were the first of their kind to offer hemophilia A & B patients a less burdensome treatment regimen (Nature Biotechnology 2014). Since then, several other extended half-life products have been approved for hemophilia A & B. Generally, the pricing of standard half-life factor products has not been reduced since the introduction of extended half-life products. In fact, the cost growth of some products has outpaced the rate of inflation when comparing 2024 Medicare ASP to inflation-adjusted 2014 Medicare ASP.

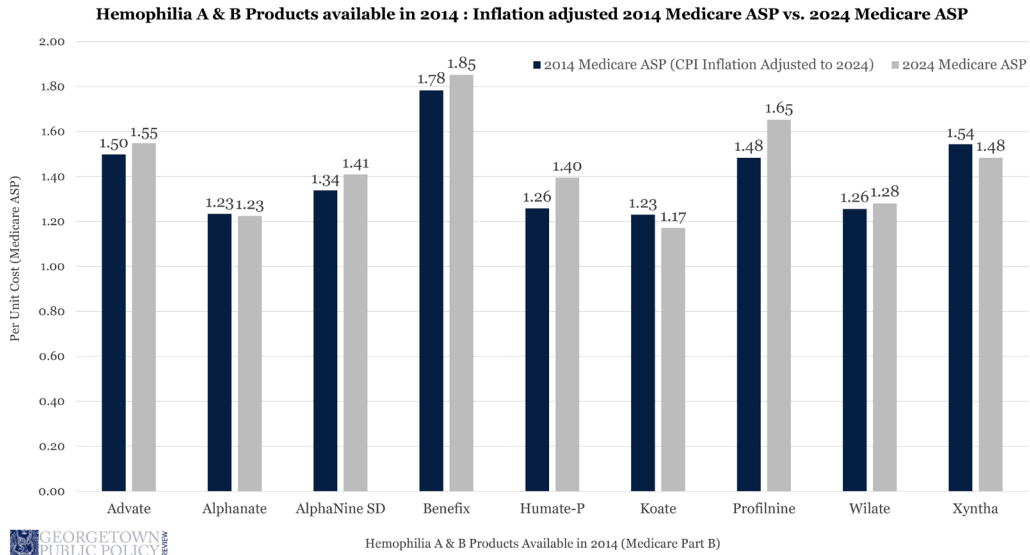


Figure 3: Standard Half-Life Factor Pricing Before and After Implementation of Extended Half-Life Factors

*EHL products approved in 2014 were not covered by Medicare until 2015. Only products covered by Medicare in 2014 are pictured.

*2014 Medicare Average Sales Price (ASP) adjusted to 2024 dollars using consumer price index (CPI)

Gene Therapy in Hemophilia

If new and improved extended half-life products have not disrupted trends in costs, which innovative products will introduce competitive pricing? In recent years, one-time gene therapy treatments have emerged as a new option for patients with hemophilia A and B. These products, CSL Behring’s Hemgenix and BioMarin’s (now-discontinued) Roctavian, were initially introduced in the US market with prices of \$3.5 million and \$2.9 million, respectively, making them the most expensive drugs in the world at that time (Myshko 2023). Though the durability of benefit is still being determined, assumptions that gene therapy benefits are long-lasting, along with high costs for standard of care treatment for hemophilia, make gene therapy “cost effective,” though not widely affordable for individual payers (i.e. insurance companies) or health systems.

Gene therapy treatments have had lower-than-anticipated uptake among patients, likely due to high per-patient costs, durability concerns, administrative challenges, and other factors. Long-term data about the durability of gene therapies for hemophilia A and B are still being collected, but some patients who

received gene therapy have had to return to factor prophylaxis. (Bala and Thornburg 2025). Due to complex logistical and administrative challenges navigating insurance coverage for gene therapy, some patients might be discouraged from pursuing treatment in the first place (World Federation of Hemophilia 2025). Finally, patients who have their bleeding well controlled using existing treatments may want to wait for a future treatment that is more likely to offer a cure for hemophilia. Slow uptake of hemophilia gene therapies has been challenging for manufacturers and has caused manufacturers to stop production. In 2025, Pfizer discontinued its gene therapy treatment for hemophilia B, Beqvez (Liu 2025) and in 2026 BioMarin discontinued its gene therapy treatment for hemophilia A, Roctavian (Becker 2025).

Consequences of High Prices

There are several important concerns related to the prices and overall spending on hemophilia treatment. First, manufacturers tend to set treatment prices high enough to capture all potential cost offsets for using a newer product, mirroring prices of

previously approved treatments. For example, a manufacturer whose product results in less frequent dosing will increase their prices to capture potential savings. This is most evident from prices set for gene therapies where forgone medication use can result in multimillion-dollar treatments being deemed cost-effective (Institute for Clinical and Economic Review 2022). However, establishing cost effectiveness does not necessarily result in improved coverage or access for patients.

Rather, high spending by payers may translate to higher cost-sharing or reduced access for patients (National Bleeding Disorders Foundation n.d.). This could be especially burdensome for patients as they face multiple insurance coverage transitions as they age. For example, patients may face little out-of-pocket spending if they are covered under a state Medicaid program, but their required out-of-pocket spending may increase as they transition to the private or employer-sponsored insurance market. In addition, patients may find that transitioning between insurers may limit the opportunity to use, or to continue to use, a preferred product (National Bleeding Disorders Foundation n.d.).

Payers may design formularies based on the per-unit costs of a drug, but this does not give a full picture of the true cost of treating someone with hemophilia. This makes it particularly important for payers to consider the total costs of care for a treatment rather than the price per unit of treatments when determining which drugs are preferred and which are not when designing formularies. Payers should consider the number of infusions required per year, the protective effect of bleed prevention, and the mitigation of further consequences such as joint replacements, mental health effects, and a costly immune resistance known as an “inhibitor” that patients may experience when they switch to a treatment that does not work for them (Gouw et al. 2013).

Barriers to Access

High prices for hemophilia products also contribute to barriers of access for patients when selecting insurance plans. Hemophilia drugs are commonly carved out by self-funded employer health plans through cost-cutting techniques like copay maximizers and alternative funding programs (AFPs) (Appleby 2022). An AFP is a

strategy intended to save money for self-funded health plans by dropping coverage for specialty drugs and attempting to find “alternatives” to traditional insurance coverage of the drug. This often involves hired vendors applying to charitable patient assistance programs on a patient’s behalf or requiring drugs to be imported from international pharmacies. AFPs don’t always achieve their intended outcome and can create unacceptable delays in medication access for patients (Gliadkovskaya 2025). This practice raises significant legal and ethical concerns.

Additionally, payers are increasingly excluding hemophilia treatments from their formularies in commercial plans (Fields n.d.). In 2026, many insurers on the Affordable Care Act (ACA) exchange have reduced their formulary to cover only one, or even zero options for hemophilia treatment (Ambetter Health n.d.). These exclusions are partly enabled by a loophole in United States Pharmacopeia (USP) classification that groups all forms of hemophilia treatment, including hemophilia A, B, and hemophilia with inhibitors, into one class (Hemophilia Alliance 2024). This is problematic because treatments are not interchangeable across types of hemophilia. Yet, it remains legal for payers to design purposefully narrow formularies that exclude most or all treatments for different types of hemophilia.

The way payers respond to high treatment costs has made it increasingly challenging for hemophilia patients to access treatment across all forms of insurance. These coverage restrictions serve as a lever for payers to reduce their overall spending on treatments but prove harmful to patients who have been able to manage their condition effectively.

Limitations

There are several limitations to our analysis. First, bleeds vary greatly in severity and the number of infusions that are required for a patient to heal. Estimating bleeding risk is difficult because each patient is unique and has a different response to treatment.

Second, individual patients may metabolize hemophilia treatments differently. Because of this, hematologists may prescribe treatment

on a different dosing frequency than what the manufacturer recommends.

Third, most hemophilia patients are covered through commercial insurance or Medicaid. Medicare data may not reflect what the average treatment for an individual hemophilia patient may cost. Due to unique factors, including the devastating history of HIV/AIDS in the hemophilia community, a significant portion of an entire generation was tragically lost. Therefore, many patients may not be of Medicare age. (White 2010).

Fourth, the figures above notably exclude the latest approvals for hemophilia, (Hympavzi, Alhemo, Qftlia). At the time of this analysis, Medicare ASP data was not readily available.

Fifth, the market for hemophilia gene therapy has drastically changed through the course of this analysis. As of March 2026, Hemgenix (for hemophilia B) remains the sole gene therapy commercially available. With one-time, multimillion dollar gene therapies falling out of the picture for the future of treatment, it remains unclear how the pricing of treatments will evolve.

Sixth, this analysis did not include every factor product marketed at the time of this analysis. Rather, the figures featured in this article mainly include treatments indicated for prophylaxis. Before prophylaxis became the standard of care for treating severe hemophilia, many previously approved factor products did not carry an indication for prophylaxis on their FDA labels. In practice, physicians may still prescribe these treatments prophylactically off label, but the lack of an indication made it difficult to find data such as median annual bleed rate for products such as Koate, Profilnine, or Alphanate.

Finally, Hemlibra is an outlier among hemophilia treatments because its non-factor mechanism of action requires a one-time loading dose. Additionally, traditional factor replacement must be used to treat bleeds while patients are on Hemlibra. The price pictured in the analysis is for the first year of treatment on Hemlibra, including loading doses and a median annual bleed rate of 0.0 (see Appendix). In a multiyear analysis, Hemlibra pricing would be slightly lower.

Conclusion

Despite the benefits of new hemophilia treatments, it is important to consider whether the current benchmark for hemophilia drug pricing is appropriate in the US. As products continue to improve, should the total benefit of foregone spending on complications avoided accrue only to drug manufacturers? Or should improvements in products contribute to meaningful cost reductions across this therapeutic class? Hemophilia treatment prices are on a troublesome path and should likely be reduced to maintain long-term sustainability and access for patients as the demand for treatments continues to increase.

The high price of hemophilia drugs is both a gift and a curse to individuals living with hemophilia. Partially due to the profitability of treatments, the pharmaceutical industry has continuously pursued research and development for bleeding disorders. These investments have resulted in improved life expectancy, reduced treatment burden, and drastically improved quality of life for individuals with hemophilia. The paradigm of hemophilia treatment has shifted, and patients can experience outcomes far greater than what was ever thought possible decades ago.

However, the high price of hemophilia drugs can make access difficult. As it currently stands, health plans seek to avoid the responsibility of coverage of these costly therapies. Efforts to reduce health plan costs, such as limiting the number of covered drugs on commercial insurance plans or carving out medication coverage through alternative funding programs, generally limit patient access. With limited access to insurance coverage, hemophilia patients are left with fewer treatment options and potentially costly delays in care. This needs to change, and patients must be centered in discussions around reducing costs.

Cost reduction efforts must focus on ways to responsibly reduce spending by promoting favorable patient health outcomes, not by deterring individuals from coverage. Payers of all types should take greater responsibility in recognizing the medical necessity of these treatments and the right for hemophilia patients to live a healthy life. If a patient is stable on a particular treatment and has demonstrated success, continued access and adherence

can result in mutually beneficial outcomes for patients and payers if complications are avoided. Otherwise, costs are simply being shifted elsewhere and not meaningfully reduced. Additionally, manufacturers should take responsibility by adjusting their pricing to reflect the current standard of care, and not the high benchmark that has been established over time. It is complex and costly to develop these treatments, but expectations for life with hemophilia are continuously evolving for the better. In an era with over a dozen drugs to choose from, it should be considered outdated for pricing to reflect an era where nonstop hospitalizations were standard for patients trying to manage their condition. The US hemophilia market features many competitors, but it needs competition.

Therefore, patients must always come first, and payers and manufacturers must take greater responsibility in easing the burden placed on patients with hemophilia. It is time to reevaluate the balance between pricing and coverage to ensure that the benefits of novel therapies can be accessed by all patients.

References

- Ambetter Health. “2026 Tennessee Formulary Changes.” Accessed March 31, 2026. <https://www.ambetterhealth.com/content/dam/centene/ambettertn/pdfs/2026-tn-formulary-change.pdf>.
- Appleby, Julie. “Patients Could Feel Pinch as Health Plans Seek to Offload Pricey Prescription Costs.” CBS News. Published December 5, 2022. Accessed March 31, 2026. <https://www.cbsnews.com/news/prescription-drug-costs-health-insurance-coverage-benefits/>.
- Bala, Natasha S., and Courtney D. Thornburg. “Gene Therapy in Hemophilia A: Achievements, Challenges, and Perspectives.” *Seminars in Thrombosis and Hemostasis* 51, no. 1 (2025): 28–40. Published online April 8, 2024. <https://doi.org/10.1055/s-0044-1785483>.
- Becker, Zoey. “BioMarin Resigned to Ending Journey with Hemophilia A Gene Therapy Roctavian, Seeks Divestiture amid Sales Struggle.” FiercePharma. Published October 28, 2025. Accessed March 31, 2026. <https://www.fiercepharma.com/pharma/biomarin-resigned-ending-journey-hemophilia-gene-therapy-roctavian-seeking-divestiture-amid>.
- “Blood Clotting Factor.” ScienceDirect. Elsevier. Accessed March 31, 2026. <https://www.sciencedirect.com/topics/medicine-and-dentistry/blood-clotting-factor>.
- Centers for Disease Control and Prevention. “About Hemophilia.” Accessed March 31, 2026. <https://www.cdc.gov/hemophilia/about/index.html>.
- CMI. “U.S. Hemophilia Treatment Market Size to Worth US\$ 6.81 Billion by 2031, Growing at a CAGR of 6.2%.” Yahoo Finance. Published July 26, 2024. Accessed March 31, 2026. <https://finance.yahoo.com/news/u-hemophilia-treatment-market-size-090700684.html>.
- Contreras, Briana. “Switching to Extended Half-Life Therapies Improves Hemophilia Care, but Raises Short-Term Costs.” Managed Healthcare Executive. Published July 18, 2025. Accessed March 31, 2026. <https://www.managedhealthcareexecutive.com/view/switching-to-extended-half-life-therapies-improves-hemophilia-care-but-raises-short-term-costs>.
- Fields, Lisa. “Coverage Denied: The Growing Problem of Formulary Restrictions.” HemAware. Accessed March 31, 2026. <https://hemaware.org/increase-of-formulary-restrictions-is-problem>.
- Gliadkovskaya, Anastassia. “A New Wave of Middlemen Offers ‘Alternative Funding’ for Specialty Drugs. Patients Bear the Risks.” Fierce Healthcare. Published October 14, 2025. Accessed March 31, 2026. <https://www.fiercehealthcare.com/payers/new-wave-middlemen-promise-savings-specialty-drugs-patients-bear-risks>.
- Gouw, Samantha C., Johanna G. van der Bom, Rolf Ljung, Carmen Escuriola, Ana R. Cid, Ségolène Claeysens-Donadel, Christel van Geet, et al. “Factor VIII Products and Inhibitor Development in Severe Hemophilia A.” *New England Journal of Medicine* 368, no. 3 (2013): 231–239. <https://doi.org/10.1056/NEJMoa1208024>.
- Healthline. “Hemophilia Life Expectancy and Primary Causes of Death.” Published August 14, 2023. Accessed March 31, 2026. <https://www.healthline.com/health/hemophilia-a/hemophilia-life-expectancy#without-treatment>.
- Hemophilia Alliance. “November 2024 Newsletter.” Published November 2024. Accessed March 31, 2026. <http://hemoalliance.org/november-2024-newsletter/>.
- Institute for Clinical and Economic Review. “ICER Publishes Final Evidence Report on Gene Therapies for Hemophilia A and B.” Published December 22, 2022. Accessed March 31, 2026. <https://icer.org/news-insights/press-releases/icer-publishes-final-evidence-report-on-gene-therapies-for-hemophilia-a-and-b/>.

- Kaiser Health News. "States Strive to Curb Costs for Crucial Hemophilia Treatment." NBC News. Published February 28, 2018. Accessed March 31, 2026. <https://www.nbcnews.com/health/health-news/states-strive-curb-costs-crucial-hemophilia-treatment-n851991>.
- Liu, Angus. "Pfizer Empties Gene Therapy Portfolio with Hemophilia Discontinuation." FiercePharma. Published February 21, 2025. Accessed March 31, 2026. <https://www.fiercepharma.com/pharma/pfizer-empties-gene-therapy-portfolio-discontinues-hemophilia-treatment-beqvez>.
- Miller, James. "Haemophilia Drugs: Why Are They So Expensive?" Liv Hospital. Published February 25, 2026. Accessed March 31, 2026. <https://int.livhospital.com/haemophilia-drugs-why-are-they-so-expensive/>.
- Myshko, Denise. "FDA Approves \$2.9 Million Gene Therapy for Hemophilia A." Managed Healthcare Executive. Published June 29, 2023. <https://www.managedhealthcareexecutive.com/view/fda-approves-2-9-million-gene-therapy-for-hemophilia-a>.
- National Bleeding Disorders Foundation. "Access to Care." Accessed March 31, 2026. <https://www.bleeding.org/advocacy/federal-priorities/access-to-care>.
- National Bleeding Disorders Foundation. "Hemophilia A." Accessed March 31, 2026. <https://www.bleeding.org/bleeding-disorders-a-z/types/hemophilia-a>.
- National Bleeding Disorders Foundation. "MASAC Document 288: MASAC Resolution on Proposed Hemophilia Treatment Guidelines from the International Society on Thrombosis and Haemostasis (ISTH)." Accessed March 31, 2026. <https://www.bleeding.org/healthcare-professionals/guidelines-on-care/masac-documents/masac-document-288-masac-resolution-on-proposed-hemophilia-treatment-guidelines-from-the-international-society-on-thrombosis-and-haemostasis-isth>.
- National Bleeding Disorders Foundation. "Patient Out-of-Pocket Expenses." Accessed March 31, 2026. <https://www.bleeding.org/advocacy/state-priorities/patient-out-of-pocket-expenses>.
- National Bleeding Disorders Foundation. "Products Licensed in the US." Accessed March 31, 2026. <https://www.bleeding.org/healthcare-professionals/guidelines-on-care/products-licensed-in-the-us>.
- Nature Biotechnology. "Long-Life Agent for Hemophilia A." *Nature Biotechnology* 32 (2014): 704. <https://doi.org/10.1038/nbt0814-704c>.
- Van Thillo, Quentin, and Cédric Hermans. "Rebalancing Agents in Hemophilia: Knowns, Unknowns, and Uncertainties." *Haematologica* 110, no. 12 (2025). <https://doi.org/10.3324/haematol.2025.288245>.
- White, Gilbert C., II. "Hemophilia: An Amazing 35-Year Journey from the Depths of HIV to the Threshold of Cure." *Transactions of the American Clinical and Climatological Association* 121 (2010): 61–73. <https://pmc.ncbi.nlm.nih.gov/articles/PMC2917149/>.
- World Federation of Hemophilia. "Critical Juncture in Hemophilia Treatment: Global Organizations Issue Urgent Call to Action." Published May 6, 2025. Accessed March 31, 2026. <https://wfh.org/article/critical-juncture-in-hemophilia-treatment-global-organizations-issue-urgent-call-to-action/>.

Appendix

Products Indicated for Prophylactic Treatment of Hemophilia A Factor VIII Deficiency (2024 Medicare ASP)

Product	Manufacturer	Condition	Product Type	HCPCS Code (Dosage)	Medicare ASP	FDA Approval Year	Recommended Prophylactic Dosage	Recommended Prophylactic Infusions/mL/infusion	Median Annual Blood Rate (mL/kg)	Half-life Hours	Cost of Single Infusion	Weekly Prophylaxis Cost	Annual Prophylaxis Cost	Adjusted Cost
AA-TWO	Sanoofi	Hemophilia A	Standard Half-life	1.U	\$ 4,034,287.0	2013	50 U/kg	1.00	0.00	40.00	\$ 22,205.00	\$ 22,205.00	\$ 1,258,414.00	\$ 1,159,419.00
Enprocentifit	Pfizer Inc	Hemophilia A	Standard Half-life	1.U	\$ 2,195,281.0	2019	50 U/kg	1.75	1.20	21.70	\$ 14,208.00	\$ 10,021.00	\$ 537,314.00	\$ 540,429.00
Enprocentifit	Pfizer	Hemophilia A	Standard Half-life	1.U	\$ 2,440,281.0	2018	40 U/kg	2.00	1.90	19.00	\$ 3,272.00	\$ 19,244.00	\$ 364,200.00	\$ 381,054.00
ADMA-BDA	Genmab	Hemophilia A	Factor Half-life	0.5 (60)	\$ 52,830,281.0	2017	1.5 mg/kg	1.00	0.00	640.00	N/A	\$ 15,050.70	\$ 600,028.00	\$ 543,220.00
ADON (A)	S.T. Biotech	Hemophilia A	Standard Half-life	1.U	\$ 1,480,281.0	2016	50 U/kg	2.00	1.74	14.20	\$ 4,935.00	\$ 20,002.00	\$ 1,261,848.00	\$ 1,080,765.00
ADONAL (A)	Pfizer	Hemophilia A	Standard Half-life	1.U	\$ 1,477,281.0	2016	40 U/kg	2.00	2.00	14.20	\$ 5,013.00	\$ 10,007.00	\$ 175,568.00	\$ 185,760.00
ADONAL (A)	Pfizer	Hemophilia A	Standard Half-life	1.U	\$ 2,095,281.0	2015	50 U/kg	2.00	1.90	14.00	\$ 4,907.20	\$ 19,002.00	\$ 1,204,000.00	\$ 1,083,007.00
ADON (A)	Cocept	Hemophilia A	Standard Half-life	1.U	\$ 1,344,281.0	2015	40 U/kg	2.00	2.00	17.10	\$ 4,172.20	\$ 15,495.20	\$ 350,200.00	\$ 364,054.00
ADON (A)	Sanoofi	Hemophilia A	Standard Half-life	1.U	\$ 2,344,281.0	2014	50 U/kg	1.75	1.60	19.70	\$ 1,134.00	\$ 11,094.00	\$ 1,111,000.00	\$ 1,011,000.00
ADON (A)	Pfizer	Hemophilia A	Standard Half-life	1.U	\$ 1,329,281.0	2013	50 U/kg	2.00	2.00	12.00	\$ 6,540.00	\$ 10,000.00	\$ 1,201,000.00	\$ 1,040,200.00
ADON (A)	Cocept	Hemophilia A, von Willebrand Disease	Standard Half-life	1.UU, VWP (65.0)	\$ 1,280,282.0	2020	40 U/kg	2.00	0.00	11.41	\$ 4,954.00	\$ 14,902.00	\$ 258,200.00	\$ 190,704.00
ADON (A)	Pfizer	Hemophilia A	Standard Half-life	1.U	\$ 1,485,282.0	2020	50 U/kg	2.00	1.90	11.60	\$ 4,202.00	\$ 12,022.00	\$ 350,200.00	\$ 301,222.00
ADONAL (A)	Pfizer	Hemophilia A	Standard Half-life	1.U	\$ 1,540,282.0	2020	40 U/kg	2.00	2.00	12.00	\$ 5,800.00	\$ 20,000.00	\$ 1,310,000.00	\$ 1,040,200.00

Products Indicated for Prophylactic Treatment of Hemophilia B Factor IX Deficiency (2024 Medicare ASP)

Product	Manufacturer	Condition	Product Type	HCPCS Code (Dosage)	Medicare ASP	FDA Approval Year	Recommended Prophylactic Dosage	Recommended Prophylactic Infusions/mL/infusion	Median Annual Blood Rate (mL/kg)	Half-life Hours	Cost of Single Infusion	Weekly Prophylaxis Cost	Annual Prophylaxis Cost	Adjusted Cost
ADONAL (B)	Pfizer Inc	Hemophilia B	Standard Half-life	1.U	\$ 4,540,281.0	2017	40 U/kg	1.00	1.00	110	\$ 17,247.00	\$ 17,247.00	\$ 600,000.00	\$ 500,000.00
ADONAL (B)	Pfizer	Hemophilia B	Standard Half-life	1.U	\$ 5,050,281.0	2016	40 U/kg	1.00	1.4	104	\$ 14,200.00	\$ 14,200.00	\$ 600,470.00	\$ 1,000,185.00
ADONAL (B)	Pfizer	Hemophilia B	Standard Half-life	1.U	\$ 1,800,281.0	2015	50 U/kg	2.00	1.52	24	\$ 1,000.00	\$ 24,000.00	\$ 1,201,000.00	\$ 1,200,200.00
ADONAL (B)	Sanoofi	Hemophilia B	Standard Half-life	1.U	\$ 3,011,281.0	2014	50 U/kg	1.00	1.4	85.00	\$ 17,007.20	\$ 17,007.20	\$ 600,000.00	\$ 1,011,000.00
ADONAL (B)	Pfizer	Hemophilia B	Standard Half-life	1.U	\$ 1,652,189.0	1997	100 U/kg	2.00	2	23.1	\$ 9,541.00	\$ 19,082.00	\$ 600,000.00	\$ 500,000.00

Commodity Market Stabilization Strategies for Latin America and the Caribbean

Stefan Doyen¹

Abstract

Commodity dependence is a major risk for the stable development of the economies of Latin America and the Caribbean (LAC). This paper argues that commodity-dependent LAC economies can reliably mitigate price or production shock vulnerabilities through a layered toolkit of hedging instruments, sovereign wealth funds, countercyclical fiscal rules, and climate finance, and that the experiences of Mexico, Chile, Colombia, Uruguay, Guatemala, and Jamaica offer a replicable model for the broader region. The concept of the resource curse or “Dutch Disease” can be understood as not a permanent constraint but a manageable obstacle on the path to sustained economic development. Mexico has been able to purchase put options on its own commodities market for oil and gas to prevent major economic losses during price shocks. Chile has been able to avoid incurring massive public debts during economic contractions caused by commodity price dips through sovereign wealth funds whose deposits are reaped from fiscal surpluses. Since 1999, Colombia has enacted progressively stricter countercyclical fiscal policy that has allowed for the Central Government’s effective responses to several crises in the span of little more than a decade. Uruguay, Guatemala, and Jamaica have been able to utilize parametric insurance to stabilize their commodity export markets and limit the economic externalities from severe weather events.

Commodity Hedging	Sovereign Wealth Funds	Countercyclical Fiscal Policy	Parametric Insurance
(Mexico) By shorting their nation’s own oil export market year over year, Mexico can hedge against collapses in demand.	(Chile) By collecting and saving a portion of the profit of state-owned enterprises, Chile can cover budgetary shortfalls during bear markets.	(Colombia) Through countercyclical fiscal policy, Colombia has been able to weather severe macroeconomic headwinds that commodity export dependent nations are particularly vulnerable to.	(Uruguay, Guatemala, and Jamaica) Through taking out insurance against extreme weather events, these nations can protect against collapses in supply in the sectors of agriculture, energy, and tourism.

Latin America and the Caribbean as Commodity Exporters

The economies of LAC are characterized by being major exporters of commodities. According to the United Nations Trade and Development, “14 out of 33 countries in the LAC region are commodity dependent, meaning commodities account for 60% or more of their total merchandise export revenue.” Further, another seven LAC countries are just below that threshold with commodities accounting for greater than 50% of total merchandise export

revenue (United Nations 2021). To illustrate, according to the World Economic Forum, Chile’s copper production accounts for 27% of the world’s copper production (Venditti 2022). Between the years 1996-2016, copper mining accounted for about 10% of the Chilean GDP. For additional context, between 1996-2016, the share of Chile’s economy attributable to copper extraction fluctuated between 5% and 20%, slowly declining to about 10% since 2006 (Copper Alliance 2018).

¹ Stefan Doyen is an MPP student at the McCourt School of Public Policy at Georgetown University

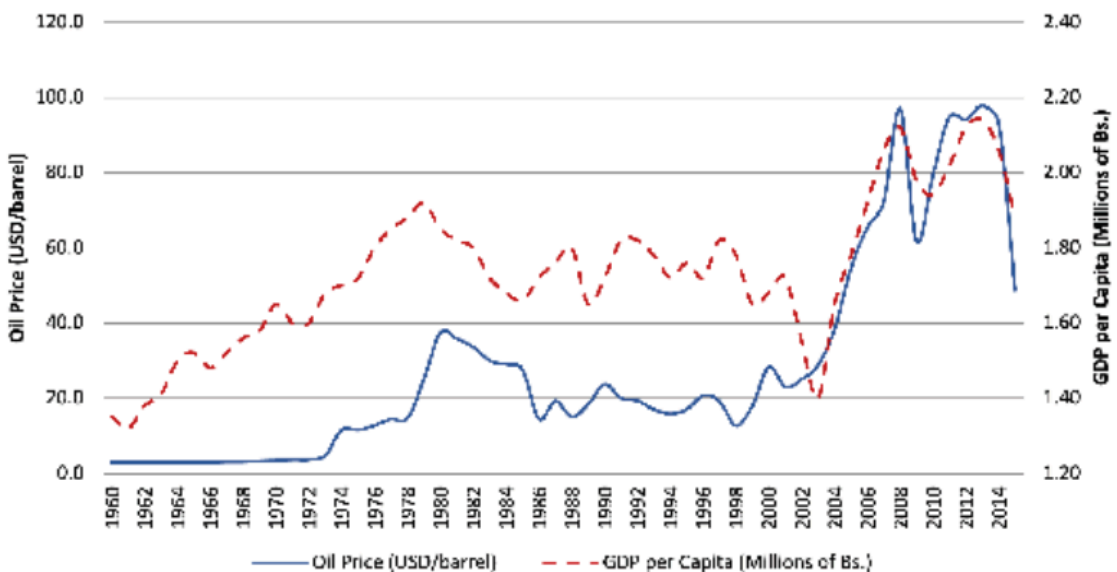
LAC economies' dependence on commodity exports can have grave consequences if there are price shocks in the market. Venezuela's experience illustrates what is at stake in the absence of these stabilization mechanisms. In 2014, the West Texas Intermediate (WTI) price per barrel of oil was over \$100 USD. By early 2016, the price plummeted to less than \$30 USD per barrel, see Chart 1.1 (Restuccia 2021). According to the Council on Foreign Relations, "Venezuela's gross domestic product (GDP) shrank by roughly three-quarters between 2014 and 2021" (2024). The economic devastation has been accompanied by widespread starvation, a mass population exodus, and a descent into increasingly dictatorial rule. However, this unmitigated economic disaster has begun to show signs of limited recovery with the Venezuelan GDP growing by 5% in 2023 (Armas 2024). Despite the recent economic stabilization, the lasting effects of the crash on Venezuela's population, democratic institutions, and long-term economic prospects will likely not revert to pre-oil-shock levels for decades (Economic Policy Institute 2009).

The Solow Growth Model holds that economic growth is a function of labor, capital, and technological progress. As such, the population exodus and capital flight caused by severe economic crashes experienced in

Venezuela will curtail its growth trajectory. The mass migration out of Venezuela will impact the nation's labor supply for many years to come, throttling economic productivity. An unstable investing environment reduces the amount of capital recirculated into the Venezuelan economy, allowing capital depreciation to continue unabated. This is evidenced by Venezuela's marked decrease in oil production since 2014 (Ycharts 2024). Venezuela's political instability will limit its ability to advance technologically, a significant loss in the nation's human capital as high-skilled workers continuously leave the country. Lastly, Venezuela's economic growth potential has been near-irreversibly eroded. It may take decades for Venezuela to regain the economic ground lost since 2014 to begin achieving earlier growth rates.

How each country in LAC enacts commodity market stabilizing policies is going to be tailored to its own institutional, political, and economic context. When considering Venezuela, there is no room for ambiguity about the consequences of eschewing any mechanism of commodity market stabilization. The following sections examine other nations in LAC that took a different path.

Venezuela: Oil Price and GDP per Capita (Chart 1.1)



Policy Solution: Commodity Hedging (Mexico)

Given the severe economic consequences of the oil price shocks of 2014-2016, it is a paramount concern of policymakers of oil-endowed nations to mitigate the effects of price shocks. Consider Mexico, the second largest oil exporter in LAC, exporting 429,192,000 barrels annually of crude oil and oil products (US Energy Information Administration 2025). Mexican oil exports account for roughly a fifth of the total oil exports in Latin America and the Caribbean (Statista 2024a). Since 2015, oil and gas extraction has averaged about 2% of Mexico's GDP, down from the 6% average from 2007-2015 (Statista 2025b). Despite the commodity cooldown triggered by the 2008 financial crisis and the collapse of oil demand during the COVID-19 pandemic, Mexico has been able to enjoy economic stability in this sector by guaranteeing minimum oil revenues through options contracts.

These options contracts are commonly referred to as the "Hacienda Hedge." Mexico's Secretary of Finance and Public Credit (SHCP) purchases options contracts with myriad investment banks (JP Morgan Chase & Co., Goldman Sachs Group Inc., Barclays, etc.) to guarantee a minimum annual revenue for Mexican oil. In 2024, the guaranteed minimum price for Mexican oil was negotiated to be \$68.70 USD/barrel (WTI). Mexico pays annual premiums to these investment banks to hold the options contracts. The premiums paid by the SHCP average around 0.1% of Mexico's annual GDP, although exact figures are not known (Valencia 2018). If the price of WTI falls below the strike price in the options contract, the Mexican government is entitled to a payout determined by the disparity between the strike price and the price of WTI.

All available data suggests the Hacienda Hedge has not only been successful at preventing losses in the Mexican oil market, but it has also been profitable at times. While the exact figure is not known, the best estimates suggest a total of \$15.1 billion USD have been paid in premiums over the twenty years on these options contracts (Blas 2017). Mexico has received a payout for four fiscal years: 2009, 2015, 2016, and 2020. According to Bloomberg, "From 2001 to 2017, the country made a profit of \$2.4 billion"; this data does not account for the premiums paid

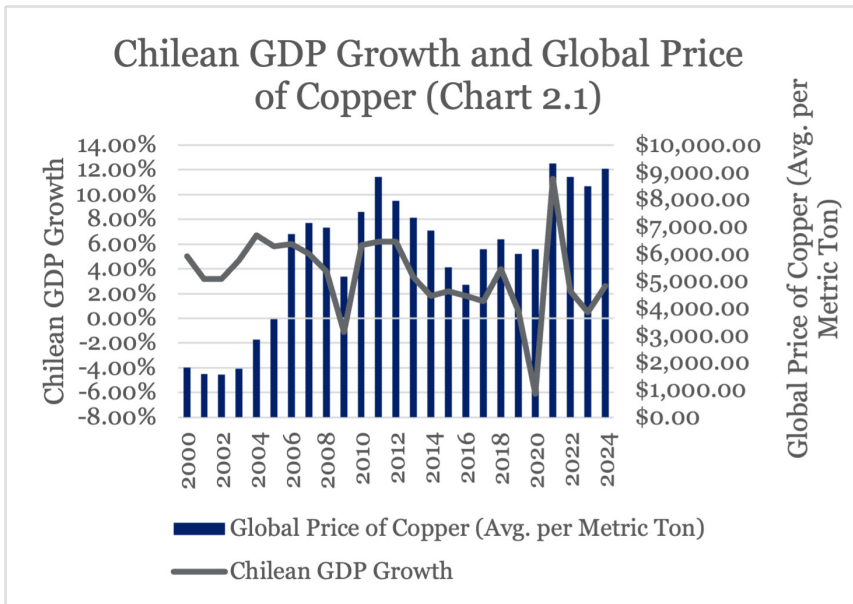
after 2017 or the payout made in 2020 (2017). Based on all available information, it appears the Hacienda Hedge generates a profit or at least breaks even.

Mexico has been able to avoid both major oil price shocks of the past twenty years. The brilliance of the program is obvious: if Mexico pays the premiums and does not receive the buyouts, it is because the WTI is higher than expected and Mexico can enjoy greater than expected revenues from oil and gas production. However, should an unforeseen price shock occur, Mexico is compensated for the deficits in expected oil revenue. In the long-term, the premiums are relatively inexpensive compared to the projected losses that could have occurred in a given fiscal year from the price shocks of 2014-2016 and 2020. As a method of mitigation, these options contracts proved invaluable. Mexico can be confident its oil and gas sector will never face the prolonged national consequences of an unmitigated commodity crash, as Venezuela has.

Policy Solution: Sovereign Wealth Funds (Chile)

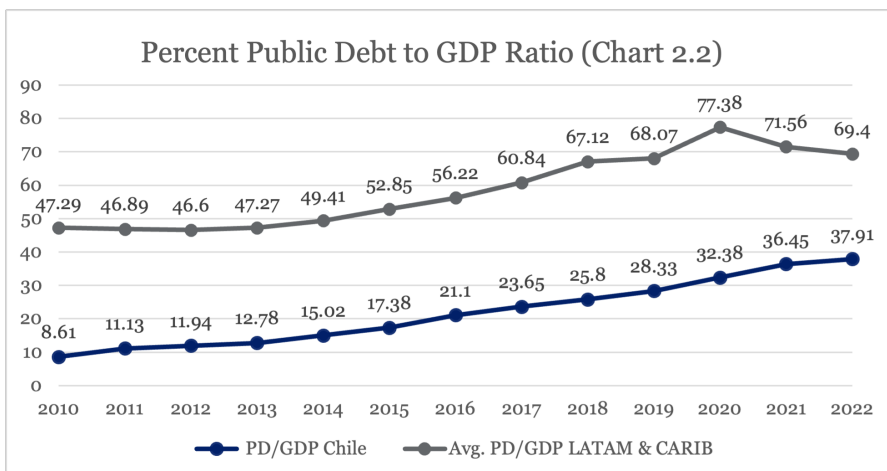
As previously stated, Chile's primary export is copper. Copper extraction constitutes roughly 10% of the Chilean economy. Analyzing data from the World Bank, we can see that from 2000-2024, Chilean GDP growth has steadily averaged 3.53% (World Bank 2024b), despite the extreme volatility of the world price for copper per metric ton since 2000 (Federal Reserve Bank of St. Louis 2025) (Chart 2.1). A critical element of Chile's success and stability has been the establishment and operation of sovereign wealth funds over almost forty years. Given the historical instability of the price of copper, it became necessary for Chile to create a "rainy day fund" to anticipate market fluctuations and limit debt issuance.

In 1987, Chile created the Copper Reserve Fund (CSF). The CSF was funded from the profits of the Chilean state-owned copper extraction firm Codelco. The purpose of this fund was to fiscally support the Central Government of Chile. In 2006, Chile enacted its Fiscal Responsibility Law, a broad piece of legislation aimed at macroeconomic stabilization. The Fiscal Responsibility Law of 2006 established the Economic and Social Stabilization Fund (ESSF),



which replaced the CSF (Ministry of Finance of Chile 2025). According to the Chilean Ministry of Finance, the purpose of the ESSF is similar to that of the CSF “to provide funding to cover fiscal deficits and/or pay back public debt.” (Ministry of Finance of Chile 2025). The key difference is the ESSF draws funds not only from the profits of Codelco but also from the government surplus. The overwhelming majority of the original balance of the ESSF was the remaining balance of the CSF.

The Fiscal Responsibility Law of 2006 accounts for the ESSF’s funding by depositing “the positive balance resulting from the difference between the effective fiscal surplus and the contributions to the Pension Reserve Fund and to the Central Bank of Chile, discounting the payment of public debt and advances made the year before” (Ministry of Finance of Chile 2025). Dividends from ESSF investments have generated only \$4.39 billion USD of the fund’s total balance, a minor share of the overall portfolio. The ratio of public debt as a percentage of GDP has steadily increased in Latin America after the commodity boom of the 2000s ended with the 2008 financial crash (Statista 2024). However, when analyzing data from the IMF, Chile has consistently had a significantly lower than average debt/GDP ratio during the same period (Statista 2025c) (Chart 2.2).



Most recently, the ESSF made major withdrawals in 2020 and 2021 to cover financial shortfalls caused by the instability of the COVID-19 pandemic. In 2020, the withdrawals totaled just over four billion USD and in 2021, the withdrawals totaled near six billion USD. The total value of the ESSF as of August 2025 is around four billion USD (Ministry of Finance of Chile 2025). Of the six largest economies in LAC, Chile was hit the fifth hardest by the COVID-19 pandemic (6% contraction), contracting slightly more than Brazil (5.8% contraction) (Cottani 2020). This can be interpreted as Chile being able to mitigate some of the worst effects of the COVID-19 pandemic compared to other large LAC economies, given that commodity-based economies with high levels of informality were among the most impacted by the pandemic. (Ohnsorge, Yu, and Kose 2022). Covering shortfalls with the ESSF has allowed Chile to avoid taking on unnecessary additional public debt. While credit risk varies country by country, reducing public deficits is important considering the average interest rate for public debt in Latin America is higher than other regions, making its financing more costly (Passadore 2024).

Policy Solution: Counter Cyclical Fiscal Policy (Colombia)

Colombia is the fourth largest economy in LAC and is highly dependent upon commodity exports, most notably oil, coal, and coffee. In 2022, commodity exports accounted for 11.5% of Colombia's economy. Following the end of the commodity boom and the oil price shocks, by 2018, Colombia had regained the second-highest Herfindahl-Hirschman Index (HHI) score for export concentration in the region, trailing only Chile. (International Monetary Fund 2023). After the ratification of the Colombian Constitution of 1991, all Colombians were now entitled to new rights including education, healthcare, and employment (Constitute Project 2015). Because of the social investment obligations required by the Colombian government, the annual fiscal deficit to GDP ratio steadily climbed. The first economic crisis to cause a major spike in the fiscal deficit was the banking crisis of 1999, raising the deficit to 6% of GDP (World Bank 2021a). However, following the banking crisis of 1999, several institutional changes were made to Colombian fiscal policy. Fiscal policy was

redirected to a strategy of countercyclicality. According to the World Bank, "during 2000–08, Colombia reduced the general government deficit and built fiscal buffers that enabled it to respond effectively to the 2008 global financial crisis" (2021b). This paper will consider three of these institutional reforms. First, Law 617 of 2000 requires fiscal sustainability of districts and municipalities. Second, Law 819 of 2003 requires the use of a Medium-Term Fiscal Framework. Third, the Fiscal Rule of 2011 establishes limits on public expenditures relative to GDP.

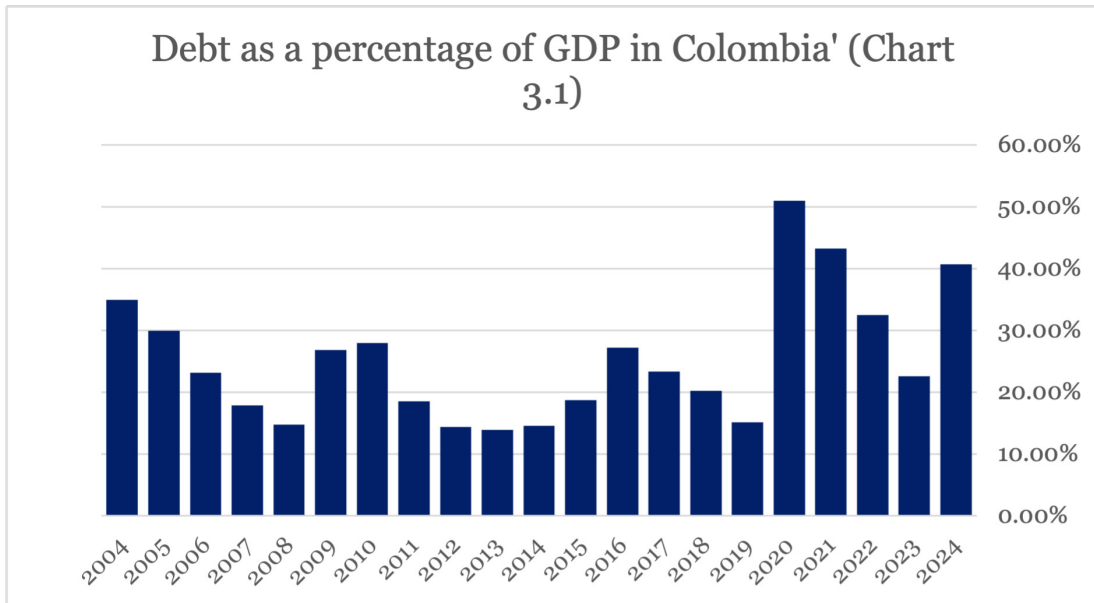
Law 617 of 2000 began by categorizing the types of municipalities in Colombia by population size and economic productivity. Special Category districts are the largest and most economically productive regions per-capita, followed by the slightly smaller first category districts, all the way down to sixth category districts. Law 617 restricts municipal spending to a specific proportion of revenue for each of the categories. Special Category is the lowest at 50%, and fourth, fifth, and sixth are the highest at 80% (Funciones Públicas de Colombia 2025). This established systemic guidelines for fiscal discipline in municipalities, mandated efficient public spending, and required transparent fiscal structures from municipalities to the central government of Colombia.

Colombia passed the Responsibility and Transparency act of 2003 (Law 819), mandating the Central Government of Colombia use a Medium-Term Fiscal Framework (MTFF). This requires the Central Government and all municipalities to annually produce a ten-year economic framework outlining balance targets. The Central Government and all municipalities were expected to strictly abide by the MTFF. The issue was that the MTFF was non-binding. Additionally, it was not possible for the MTFF to accurately account for revenue fluctuations due to the business-cycle volatility to which commodity-export-dependent economies are particularly sensitive (Vargas, González, and Lozano 2012). While the MTFF was a part of a series of successful reforms and fiscal management tools, in 2011 the weaknesses of Law 819 were corrected.

The Fiscal Rule of 2011 (Law 1473) changed the non-binding qualitative MTFF into a binding quantitative framework. The goal of Law 1473 was to reduce the public deficit to 1% of GDP by 2022. This goal was incrementally approached

from 2011 to 2022. Each year the proportion of public deficit to GDP was steadily lower, with a target of 1%. However, the massive public spending required to fight COVID-19, coming on the heels of the oil price shocks of 2014–2016, placed enormous strain on a government whose debt-to-GDP ratio was already 50% in 2019. (World Bank 2021a). Unfortunately, COVID-19 caused the worst economic crisis in Colombia's

history. Colombia marshalled the remainder of its fiscal room to combat COVID-19 medically and economically, driving “the debt-to-GDP ratio to unprecedented levels” (World Bank 2021a). The lesson to learn is that Colombia was able to mobilize significant fiscal resources to deal with these crises because of earlier, prudent countercyclical policy.



Despite profound macroeconomic instability affecting Colombia's prospects for economic growth, Colombia has been able to shoulder the worst effects of these headwinds through responsible fiscal policy frameworks. Data from Colombia's Ministry of Finance and Public Credit illustrate these trends clearly: spending spikes following the 2008 Financial Crisis, Oil Price Shocks 2014–2016, and COVID-19 followed by spending regressions (Chart 3.1) (Central Bank of Colombia 2025). Immediately following the Financial Crisis of 2008, Colombia was still able to achieve an improved rating of “Investment Grade” from Standard & Poor's in 2011 (Stott and Schipani 2025). However, the profound impact of the COVID-19 pandemic caused Colombia to lose that rating in 2021

but partially regain the rating according to some credit rating agencies in 2025. Moody's cited Colombia's downgrade because of the recent suspension of the Fiscal Rule of 2011 (Yapur 2025). Whether Colombia regains ‘Investment Grade’ will depend on its recommitment to countercyclical fiscal policy.

Policy Solution: Climate Finance (Parametric Insurance)

While the preceding sections address price volatility in traditional commodity markets, LAC economies face an equally destabilizing class of shocks: climate-driven collapses in commodity supply. For nations whose primary

export is tourism, agricultural output, or renewable energy, a hurricane or drought is functionally equivalent to an oil price crash, an externally caused revenue shock. Climate finance, or more specifically, parametric insurance addresses this gap. Climate finance encompasses financial instruments designed to mitigate the economic consequences of climate-driven shocks. A common financial instrument to mitigate climate risk and stabilize supply-side commodity production is parametric insurance. Parametric insurance disburses a pre-negotiated payout when a “parameter” is met. A parameter can be any agreed upon measurable event, like a drought, hurricane, or a threshold of seismic activity. Nations across LAC can and do utilize parametric insurance for myriad reasons. A climate catastrophe can cause irreparable financial harm through damaged facilities and shortened tourist seasons to small Caribbean nations dependent on tourism. Low wind speeds in a nation heavily invested in wind power can lead to energy shortages. To an economy heavily dependent on agricultural exports, a season of low or excessive rainfall can dramatically reduce total exports. By providing rapid lifelines of liquidity, parametric insurance can instantly respond to any kind of climate shock, avoiding the traditional time-consuming debt-seeking process after these disasters. In contrast to the Hacienda Hedge, a demand-side price stabilization tool, parametric insurance directly addresses the supply-sides of commodities markets. It remains a promising, yet underutilized tool that commodity-export dependent nations can use to hedge against production shocks.

Attractive geography is as much a natural endowment as copper and crude oil. Since tourism is classified as an export, this paper treats it as a commodity export (US Dept. of Commerce 2025). When tourism is considered a commodity export, the role of climate insurance in smaller Caribbean nations becomes apparent. Tourism constitutes one fifth to one third of the Jamaican economy (Robles 2025). Unfortunately, the nation is geographically vulnerable to extreme weather events that have disrupted lucrative tourist seasons and damaged critical infrastructure. During late October and early November of 2025, Hurricane Melissa made landfall in Jamaica, causing an estimated 8.8 billion dollars in damages (World Bank and Inter-American Development Bank 2025). In 2024, Jamaica re-

newed a catastrophic (cat) bond with the World Bank. A cat bond is a type of parametric insurance for extreme storm events. After Hurricane Melissa, Jamaica received the maximum payout for the cat bond totaling a sum of 150 million dollars (World Bank 2025). While a 150-million-dollar payout is relatively small compared to the damages, it represents an immediate cash influx for a nation facing a desperate situation. Notwithstanding an uncommonly strong storm, this is a promising model that can be applied to many Caribbean nations that are vulnerable to extreme weather events.

Energy extraction can be throttled by extreme climate events. Oil refineries are not designed to operate under conditions of extreme cold or extreme heat. In February 2021, frigid temperatures across the US reduced production of oil by “nearly 4 million barrels per day” (DTN 2023). Climate volatility can have a significant effect on a nation’s energy output. Shifting the focus back to LAC, Uruguay has become a model nation for renewable energy sourcing across LAC. Currently, about 99% of Uruguay’s energy portfolio is derived from renewable sources. Only 1-3% of electricity is generated from thermal plants that use fossil fuels. Meanwhile, “hydropower accounts for 45%, wind can contribute up to 35% of total electricity, and biomass ... accounts for 15%. Solar fills in the gaps” (Silverstein 2025). Despite Uruguay’s diversified renewable energy portfolio, the lion’s share of its renewables, hydroelectric and wind, are vulnerable to climate irregularities. If there is a drought or a windless period, Uruguay would need to rely more heavily on its thermal plants. In 2012, a drought reduced hydroelectric output and forced Uruguay to purchase 1.4 billion dollars’ worth of oil to cover its unanticipated energy shortfalls. Uruguay’s state utility company (UTE), had to pass the costs on to the consumers and receive additional funds from the government (World Bank 2013). Facilitated by the World Bank, in 2013 and 2018, UTE developed an entirely novel parametric insurance package to cover energy shortfalls caused by droughts (International Monetary Fund 2023), ensuring budgetary stability and a continuous energy supply for its citizens. As nations in LAC continue to invest more heavily in the renewable energy sector, they can use weather hedging financial instruments to protect their energy supply from the unpredictability of the environment. While

parametric insurance in the case of Uruguay is limited to energy generation and domestic consumption, the insurance mechanism is still valid for energy commodities extracted for export.

LAC is the world's leading exporter of agricultural commodities. From 2014 to 2024, agricultural exports accounted for about 6% of the regional GDP for LAC, though this can vary greatly between nations (ECLAC 2024). According to a 2025 UN report, 74% of LAC countries are highly exposed to extreme weather events that affect food security (Food and Agriculture Organization of the United Nations 2025). Either excess rainfall or droughts can decimate crop yields. In a region dependent on agricultural exports, that is a devastating proposition. To counter this, in 2021, Guatemala's Ministry of Agriculture and Livestock introduced the Catastrophic Parametric Insurance (CPI) program. So far, the pilot program has been successful, currently 100,000 farmers are participating in the program across Guatemala. The CPI measures the triggering parameter by quantities of rainfall as measured by satellite (World Bank 2024a). In cases of severe weather, the farmers can receive an individual payout from a nearby ATM or bank. Unfortunately, the climate insurance industry for agricultural shocks, while promising, has had very little penetration in LAC. However, the Inter-American Institute for Cooperation on Agriculture is working on expanding parametric insurance for farmers in Southern Cone countries (IICA 2025). Despite parametric insurance markets having little penetration in LAC agricultural commodities markets, it remains a promising mechanism for stabilizing export revenues for LAC nations that export agricultural commodities.

Parametric insurance represents a pragmatic solution for how LAC nations can develop resilient policies that function as a hedging instrument to maintain stable commodity export revenues. As a strategic tool, parametric insurance supports rapid disaster response in Jamaica, delivers energy security in Uruguay, and provides protection for rural farmers in Guatemala. Smaller and more vulnerable economies stand to benefit the most from the safety nets provided by parametric insurance. The challenge is to expand these programs across LAC nations and increase the amount of insurance available to be more effective for larger environmental disasters.

Conclusion

Commodity dependence is persistent across LAC countries. While commodity hedging, sovereign wealth funds, countercyclical fiscal rules, and climate finance all have demonstrated immense efficacy and potential, they are underutilized. Most especially climate finance, which offers a completely distinct utility by addressing supply shocks. The individual risk management frameworks developed by any one country could be fruitfully used by many more in the region. However, climate finance is somewhat an exception to this. While still a viable instrument for larger LAC nations, it is more accessible to smaller, lower-income LAC economies that lack the institutional capacity for sovereign wealth funds or complex options markets. Optimal deployment of these risk management strategies would require them to be expanded from nation-specific contexts to a region-wide common practice. Although their integration into any nation would require careful consideration of each nation's institutional capacity and economic context, their potential persists. These instruments of stabilization policy can be used complementarily. The case of Chile makes this abundantly clear. Countercyclical fiscal policy requires financing debt during bear markets but with the support of a sovereign wealth fund that was generously endowed during a bull market, it is not as daunting a task as it may seem. Combining them into a layered risk management strategy is likely the best method to achieve consistently stable outcomes for a nation's economy.

Mexico, Chile, Colombia, Uruguay, Guatemala, and Jamaica are representative of the LAC region in their economic makeup. The rest of LAC faces the same problems as these nations; it is to be hoped the solutions prove equally effective. Unfortunately, these nations stand apart from the rest of LAC as outliers because of their ability to achieve consistently stable growth. However, the experiences of these nations make clear commodity dependence need not be a permanent constraint to growth. A growing middle-income commodity rich region like LAC should employ commodity stabilization policies as a key macro-economic tool to maintain stable growth. Adoption of country specific integrated commodity stabilization policies across LAC will require tremendous technical expertise and international support. As for the World Bank, the development and promotion of parametric

insurance markets in LAC is critical for achieving an effective level of penetration. Providing continued sober advisement of countercyclical fiscal responsibility that is uninfluenced by prevailing political winds is invaluable. Lending expertise to developing nations to help develop complex commodity hedging schemes may be what is needed to expand the practice. The World Bank's commitments to these goals can help create the institutional capacity and commitment necessary for other nations in LAC to adopt these policies. A commitment to integrating these commodity stabilization policies may be the foundation for long-term prosperity that has so far proved elusive to LAC.

References

- Armas, Mayela. "Venezuela Economy Grew 5% in 2023, Will Reach 8% This Year – Maduro." Reuters, January 15, 2024. www.reuters.com/world/americas/venezuela-economy-grew-5-2023-will-reach-8-this-year-maduro-2024-01-15/.
- Banco de la República. "Serie Histórica del Balance Fiscal del Gobierno Nacional Central." Banco de la República. <https://www.banrep.gov.co/es/estadisticas/sector-publico-y-deuda-publica/gobierno-nacional-central/serie-historica-balance-fiscal-gobierno-nacional-central>.
- Bank for International Settlements. BIS Papers No 67(i). "Macroeconomic Effects of Structural Fiscal Policy Changes in Colombia." bis.org/publ/bppdf/bispap67i.pdf.
- Blas, Javier. "Uncovering the Secret History of Wall Street's Largest Oil Trade." Bloomberg, April 4, 2017. www.bloomberg.com/news/features/2017-04-04/uncovering-the-secret-history-of-wall-street-s-largest-oil-trade?embedded-checkout=true.
- "BP243: The U.S. Economy and the Great Recession." Economic Policy Institute. See epi.org/publication/bp243/.
- Colombia. "Norma N° 3771: Subtema: Financiación ... las disposiciones de ésta norma." Funciones Públicas de Colombia. funcionpublica.gov.co/eva/gestornormativo/norma.php?i=3771#.
- *Constitution of the Republic of Colombia, 2015.* Constitute Project. constituteproject.org/constitution/Colombia2015.
- Copper Alliance. "The Impacts of Copper Mining in Chile: Economic and Social Implications for the Country". April 4, 2018. sustainablecopper.org/wp-content/uploads/2018/05/ICA-Summary-Documents-The-Impacts-of-Copper-Mining-in-Chile-FV-04.04.2018.pdf.
- Council on Foreign Relations. "Venezuela Crisis." www.cfr.org/backgrounder/venezuela-crisis.
- Cottani, Joaquín. "The Effects of Covid 19 on Latin America's Economy." Center for Strategic and International Studies (CSIS), November 18, 2020. www.csis.org/analysis/effects-covid-19-latin-americas-economy.
- "Crude Oil Exports: U.S. to Mexico (Annual)." U.S. Energy Information Administration. www.eia.gov/dnav/pet/pet_move_expc_dc_nus-nmx_mbbbl_a.htm.
- DTN. "How Extreme Weather Disrupts the Oil and Gas Sector." How Extreme Weather Disrupts the Oil and Gas Sector
- The Economic Commission for Latin America, "Value of Goods Exports in Latin America and the Caribbean Will Grow 4% in 2024, New Report by ECLAC." ECLAC, <https://www.cepal.org/en/pressreleases/value-goods-exports-latin-america-and-caribbean-will-grow-4-2024-new-report-eclac>. December 12, 2025
- Federal Reserve Bank of St. Louis. "Global Price of Copper (PCOPPUSDM)." FRED, Federal Reserve Bank of St. Louis. See <https://fred.stlouisfed.org/series/PCOPPUSDM>.
- Food and Agriculture Organization of the United Nations. "New UN Report: 74 Percent of Latin American and Caribbean Countries Are Highly Exposed to Extreme Weather Events, Affecting Food Security." January 27, 2025. <https://www.fao.org/newsroom/detail/new-un-report--74-percent-of-latin-american-and-caribbean-countries-are-highly-exposed-to-extreme-weather-events--affecting-food-security/en>
- Inter-American Institute for Cooperation on Agriculture, "The Countries That Make Up the Southern Agricultural Council (CAS) Are Working Closely with the World Bank and IICA to Create Insurance Cover for Extreme Weather Events That Affect Agricultural Production." IICA, <https://iica.int/en/press/news/the-countries-that-make-up-the-southern-agricultural-council-cas-are-working-closely-with-the-world-bank-and-iica-to-create-insurance-cover-for-extreme-weather-events-that-affect-agricultural-product/>. December 12, 2025

- International Monetary Fund. “An Overview: Composition of Colombia’s ... points higher than in 2006.” IMF Working Paper, 2023. www.elibrary.imf.org/view/journals/002/2023/121/article-A003-en.xml#.
- International Monetary Fund. “Policy Dialogue: To Hedge or to Self-insure? The Benefits of Mexico’s Oil Hedging Program.” imf.org/external/np/blog/dialogo/032618.pdf.
- International Monetary Fund, “Uruguay Is Leading the Green Transition...” IMF, 2023, <https://www.elibrary.imf.org/view/journals/002/2023/179/article-A003-en.xml>.
- “Market Value.” Hacienda – Economic and Social Stabilization Fund (Chile). www.hacienda.cl/english/work-areas/international-finance/sovereign-wealth-funds/economic-and-social-stabilization-fund/financial-situation/market-value.
- Ministry of Finance of Chile. “About the Funds.” Ministry of Finance of Chile. <https://www.hacienda.cl/english/work-areas/international-finance/sovereign-wealth-funds/about-the-funds>.
- Ministry of Finance of Chile. “Economic and Social Stabilization Fund.” Ministry of Finance of Chile. <https://www.hacienda.cl/english/work-areas/international-finance/sovereign-wealth-funds/economic-and-social-stabilization-fund>.
- Ohnsorge, Franziska, Shu Yu, and M. Ayhan Kose. “The Informal Sector: Compounding the Damage of Covid-19.” VoxEU.org, 27 Jan. 2022, <https://cepr.org/voxeu/columns/informal-sector-compounding-damage-covid-19>
- Passadore, Juan & Treviño, Juan, co-leads. “Public Debt Dynamics in Latin America: Time to Rebuild Buffers and Strengthen Fiscal Frameworks (Background Paper 2).” International Monetary Fund, October 2024.
- Restuccia, Diego. “The History of Venezuela.” In “A Monetary and Fiscal History of Latin America, 1960–2017”, University of Toronto & NBER. manifold.bfi.uchicago.edu/read/a-monetary-and-fiscal-history-of-latin-america-1960-2017/section/d2c8d793-bd56-4e90-bd22-5d07377b372a.
- Robles, Frances “More Than Half a Million Jamaicans Whose Jobs Depend Directly or Indirectly on Tourism Face an Uncertain Winter After Hurricane Melissa.” The New York Times, 1 Dec. 2025, <https://www.nytimes.com/2025/12/01/world/americas/hurricane-jamaica-tourism.html>.
- Silverstein, Ken. “Uruguay’s Renewable Charge: A Small Nation, a Big Lesson for the World.” Forbes, 19 Oct. 2025, <https://www.forbes.com/sites/kensilverstein/2025/10/19/uruguays-renewable-charge-a-small-nation-a-big-lesson-for-the-world/>.
- Statista. “Latin America: Crude Oil Production by Country 2024.” Statista, June 2025a. www.statista.com/statistics/961585/latin-america-crude-oil-production-country/.
- Statista. “Mexico: Share of GDP From Oil & Gas Sector.” Statista, February 2025b. www.statista.com/statistics/1173837/gdp-share-oil-gas-sector-mexico/.
- Statista. “National Debt of Chile in Relation to Gross Domestic Product (GDP).” Statista, April 2025c. www.statista.com/statistics/370396/national-debt-of-chile-in-relation-to-gross-domestic-product-gdp/.
- Statista. “Public Debt as Share of GDP in Latin America & the Caribbean.” Statista, April 2024. www.statista.com/statistics/1054731/public-debt-share-gdp-latin-america-caribbean/.
- United Nations Conference on Trade and Development. “The Recent Commodity Price Surge: A Boon for Latin America and the Caribbean?” UNCTAD, August 4, 2021. unctad.org/publication/recent-commodity-price-surge-boon-latin-america-and-caribbean.

- U.S. Department of Commerce. "Tourism as Trade: How Summer Travel Becomes a U.S. Export." Trade.gov. <https://www.trade.gov/feature-article/tourism-trade-how-summer-travel-becomes-us-export>
- Venditti, Bruno. "Which Countries Produce the Most Copper?" World Economic Forum, December 12, 2022. www.weforum.org/stories/2022/12/which-countries-produce-the-most-copper/.
- "Venezuela Crude Oil Production (Annual)." YCharts. ycharts.com/indicators/venezuela_crude_oil_production_annual.
- "Crude Oil Exports: U.S. to Mexico (Annual)." U.S. Energy Information Administration. www.eia.gov/dnav/pet/pet_move_expc_dc_nus-nmx_mbbbl_a.htm.
- World Bank. "Colombia Public Finance Review". 2021a, © World Bank. hdl.handle.net/10986/37168.
- World Bank. "Colombia Public Finance Review (Data Supplement)". 2021b. openknowledge.worldbank.org/server/api/core/bitstreams/4eafadf5-4e8d-511e-af3a-6b4ea154cb3a/content.
- World Bank, "Financial Protection against Climate Risks for Family Farmers in Northern Central America." World Bank March 14, 2024a. Blogs, <https://blogs.worldbank.org/en/latinamerica/financial-protection-climate-risks-family-farmers-northern-central-america>.
- World Bank. "GDP Growth (Annual %) – Chile." World Bank, 2024b. <https://data-worldbank.org/indicator/NY.GDP.MKTP.KD.ZG?contextual=default&end=2024&locations=CL&start=2001&view=chart>.
- World Bank "Hurricane Melissa Triggers 100% Payout of \$150 Million World Bank Catastrophe Bond for Jamaica." World Bank, 7 Nov. 2025, <https://www.worldbank.org/en/news/press-release/2025/11/07/hurricane-melissa-triggers-100-payout-of-150-million-world-bank-catastrophe-bond-for-jamaica>.
- World Bank and Inter-American Development Bank, "IDB Estimate Hurricane Melissa Damage to Jamaica Totals All-Time-High of US\$8.8 Billion." World Bank, 19 Nov. 2025, <https://www.worldbank.org/en/news/press-release/2025/11/19/world-bank-idb-estimate-hurricane-melissa-damage-to-jamaica-totals-all-time-high-of-us-8-8-billion>.
- World Bank, "Seguro Climático en Uruguay." World Bank, 19 Dec. 2013, <https://www.worldbank.org/en/news/feature/2013/12/19/seguro-climatico-uruguay>.
- Yapur, Nicolle. "Moody's Downgrades Colombia Citing Suspension of Fiscal Rule." Bloomberg, June 26, 2025. <https://www.bloomberg.com/news/articles/2025-06-26/moody-s-downgrades-colombia-citing-suspension-of-fiscal-rule#>.
- "Colombia Lost Its Decade-Old Investment Grade Status." *Financial Times*, June 27, 2025. <https://www.ft.com/content/320c1ae7-811d-4b7c-bf6b-9de19a2e5b70>.

Designing Effective Social Media Regulations

Grace Falgoust¹

Abstract

Consider a 15-year-old in rural Florida—let’s call her Sarah. She identifies as gay in a community with no LGBTQ+ support infrastructure. Her family relationships are strained. She spends approximately two hours daily on social media, where she has connected with an online community of LGBTQ+ youth. Research demonstrates that such online communities serve critical functions for LGBTQ+ adolescents, including identity management, peer connection, and social support (Berger et al. 2022). Florida’s House Bill 3, currently in rulemaking, would require parental consent for Sarah to have social media platforms, which they likely would not give. Sarah’s case raises a fundamental policy question: How do we design regulations that address documented harms on social media while preserving beneficial uses for populations who may depend on digital spaces? This question is increasingly urgent. Nearly one-third of US adolescents report persistent feelings of sadness or hopelessness (CDC 2021), and approximately half report experiencing online bullying (Thorn 2025). Excessive or problematic social media use has been linked to loneliness (HHS 2023a), anxiety, and depression (HHS 2023b). In response, lawmakers across the political spectrum agree that greater protections for youth are necessary. In 2024 alone, states enacted dozens of youth online safety laws, with hundreds more under consideration in 2025 (Brennen & Sanderson 2024; NCSL 2025). However, policy existence does not guarantee policy effectiveness. This paper examines the current state of social media legislation in the United States, evaluates what scientific evidence reveals about social media’s effects on adolescents, and offers recommendations for developing effective regulations.

The Current State of Social Media Legislation

Despite the centrality of social media to adolescent life, federal regulation remains limited. The Children’s Online Privacy Protection Act (COPPA) of 1998, which restricts data collection from children under 13, remains the primary federal statute governing youth online protections. Beyond COPPA, federal legislative efforts have largely stalled, including unsuccessful attempts to ban TikTok (Brennen & Sanderson 2024).

In the absence of federal leadership, states have stepped in aggressively. Since 2023, hundreds of technology-related bills addressing youth online safety have been introduced across nearly every state (Brennen & Sanderson 2024; NCSL 2025). These laws vary widely in scope. However, they can generally be grouped into six core policy tools: parental consent requirements, usage or time restrictions, content restrictions, privacy protections, addictive design regulations, and educational initiatives. Table 1 contains additional details and examples of relevant legislation. However, state efforts are constrained by federal precedent. Notably, courts have frequently ruled that broad content-based restrictions on social media access violate

First Amendment protections. In 2024, courts in Arkansas and Ohio struck down laws restricting minors’ access to social media platforms on free speech grounds (Bernard 2024).

The legal landscape shifted in June 2025, when the Supreme Court upheld Texas H.B. 1181, which requires age verification for adult websites. The Court determined that the law only “incidentally burdens protected speech” while validly serving the state interest in protecting minors (Supreme Court of the United States 2025). In plain language, the Court found that because the age-verification requirement imposes only a limited or indirect burden on free expression, it can be justified as a permissible way to advance the state’s interest in protecting the health and well-being of youth. This decision enabled the Texas law to take effect and reignited similar legislative efforts in more than 20 states, establishing important precedent for broader age-verification requirements (MultiState). While significant, the ruling does not resolve ongoing questions about the effectiveness, equity, or privacy implications of such systems.

1 Grace Falgoust is a Ph.D. student in the Psychology Department at Georgetown University and MPP student at the McCourt School of Public Policy.

Policy Tool	Description	Example Legislation
Parental Consent	Require parental permission for minors obtaining social media, or that ban minors from obtaining social media.	Florida HB3: Prohibits accounts for children under 13; requires parental consent for ages 13-15.
Usage Restrictions	Limit the amount of time minors can spend on social media platforms.	Virginia SB 854: Implements limits on time users can spend on apps.
Content Restrictions	Prohibit minors from accessing sexually explicit or other harmful content.	Louisiana Act 440: Requires age verification to prevent minors from accessing pornographic websites. Nebraska LB 1092: Addresses broader harmful content beyond sexually explicit material, including violence, self-harm, dangerous challenges, bullying, and predatory behavior.
Privacy Protections	Require privacy settings and data protection for minor users (like privacy-by-default settings, data minimization, or prohibitions on targeted advertising).	Maryland “Kids Code” (HB 603 & SB 571): Requires data protection impact assessments for services accessed by minors; mandates privacy-protective defaults; restricts unnecessary data processing.
Addictive Design Codes	Address platform design features that may be addictive or harmful.	Minnesota HF 2: Requires evidence-based mental health warnings. California SB 976: Prohibits “addictive feeds” and certain notifications without consent; requires parental controls and platform reporting. California SB 1504: Requires accessible tools for minors to report harassment or harmful content New York S7695A: Bans algorithmic feeds without consent; restrict nighttime notifications; enable screen-time limits.
Educational Initiatives	Provide funding for digital literacy, such as school-based instruction on social media risks, cyberbullying reporting strategies, and other personal risk-reduction skills.	North Carolina HB 959: Mandates instruction on social media health effects, potential addiction, cyberbullying reporting strategies, and personal skills to reduce online risks.

Scientific Understanding of Social Media’s Impact on Adolescents

Evaluating social media policy requires a clear-eyed look at what research reveals about its effects on young people. Despite widespread public concern, the evidence that typical social media harms use adolescent well-being is mixed and often small in magnitude. Studies consistently find small, inconsistent effects. Early influential work found that the association between social media and well-being was no larger than the association between wearing glasses and wellbeing (Orben & Przybylski 2019). Subsequent large-scale research has reinforced this finding. A meta-analysis of 226 studies covering over 275,000 participants found no significant overall effect on well-being, though it did identify modest links to higher anxiety and depression, and — in the opposite direction — improved social connection (Hancock et al. 2022). Researchers have also attempted to establish causation directly, by asking participants to stop using social media altogether. But even these experimental studies, when pooled together, produced effects no different from zero; and the more rigorously designed the study, the weaker the effects tended to be (Ferguson 2025). In short, the evidence does not support the conclusion that social media is uniformly or straightforwardly harmful to adolescents.

This does not mean social media has “no impact.” Rather, evidence suggests that how adolescents use social media matters far more than how much they use it. Systematic reviews focusing on problematic or excessive use demonstrate more consistent associations with anxiety, depression, and distress. This distinction has profound policy implications because it points to different regulatory strategies. If harms were primarily dose-dependent, “sledgehammer” policies like time limits or age-based restrictions would be appropriate. However, the evidence suggests that the most effective regulations should focus on “scalpel” approaches that target problematic and excessive use, rather than limiting use overall.

Further, evidence suggests how adolescents use social media is often facilitated or inhibited by specific platform design choices. Features such as algorithmic recommendation systems, visible metrics of social approval, encrypted

messaging, and disappearing content can amplify risks of harm, including social comparison, cyberbullying, and sexual exploitation (Thorn 2025). In contrast, other design features can reduce engagement and associated risks. When Facebook and Instagram experimentally implemented semi-chronological feeds in 2020, time spent declined by 20% on Facebook and 10% on Instagram (Guess et al. 2023).

For policymakers, this evidence base reveals both an urgent need and a genuine opportunity. The harms associated with social media are real, but importantly, they are not inevitable. Research shows that negative outcomes are often driven by specific platform design choices, indicating that regulatory interventions could mitigate these risks. The central question, then, is not whether to intervene, but how to do so effectively.

Tensions in Designing Social Media Regulations

Although lawmakers broadly agree that youth need greater online protections, their preferred approaches diverge. Democratic-led states tend to emphasize platform accountability, privacy protections, and design restrictions. Republican-led states more often prioritize parental authority through consent requirements and age verification, particularly for sexually explicit content.

These approaches reflect deeper tensions within children’s rights frameworks. Specifically, the balance between participatory rights (autonomy, expression, exploration) and protective rights (safety, privacy, freedom from harm). Regulations often struggle to satisfy both. For example, restricting sexually explicit content may reduce exposure to harmful material but can also block access to legitimate resources such as sexual health information or LGBTQ+ support forums.

Well-intentioned policies can also create exclusionary barriers. Age-verification systems may require documentation that undocumented youth or those in foster care lack. Parental consent requirements assume stable, supportive family environments—an assumption that does not hold for many vulnerable adolescents.

Finally, there is the challenge of enforcement. It is challenging to ensure platform compliance. If not all platforms

comply, there may be unintended consequences. For example, Louisiana's Act 440, implemented in 2022, required age verification for adult content.

Three months post-implementation, searches for compliant platforms declined 51%, while searches for non-compliant platforms increased 48%, and VPN searches rose 24% (Lang et al. 2025). This pattern suggests substitution toward less regulated and potentially riskier platforms. Rather than preventing access, the law appears to have redirected users to platforms with fewer safeguards and greater exposure to harm.

Further, identifying youth requires effective age verification. However, age verification systems that protect privacy while avoiding exclusionary effects are challenging to build. Evidence from existing policies highlights these trade-offs. For example, China maintains among the world's strictest controls on youth social media access. Yet research reveals 77% of surveyed minors evaded verification using relatives' identification (Zhao 2024), and over 1/3 of parents helped children bypass limits (National Press and Public Administration 2024). These findings suggest that even highly restrictive systems may struggle to achieve compliance in real-world settings. More broadly, there is little evidence that any single approach is fully effective. Self-report is often unreliable. Image-based solutions, like those tested in Australia, are still not entirely accurate (Kaye 2025). At the same time, stronger verification methods based on official identification (e.g., IDs or passports) raise privacy concerns and can create access barriers, particularly for undocumented youth or those in foster care.

Case Study: Nebraska's Legislative Bill 504

To explore these tensions, we present Nebraska's Legislative Bill 504 as a case study. The law exemplifies a design-focused, or "scalpel," approach to youth online safety. Rather than imposing blanket bans or time limits, the law targets specific platform features associated with documented harms.

Introduced by Senator Carolyn Bosn on January 21, 2025, and signed into law just four months later on May 30, 2025, LB 504 is among the most recent state-level youth online safety laws in the country, taking effect on January 1, 2026. Nebraska developed the law in response

to legal challenges faced by similar legislation in California, refining its approach to improve durability. In this sense, LB 504 reflects states acting as laboratories of democracy. It builds on earlier efforts, notably California's legislation and the UK's Age-Appropriate Design Code, while using more flexible language that gives companies discretion in how to comply. What distinguishes LB 504 is how it synthesizes those precedents into a legally cautious, implementation-ready framework well-positioned for broader adoption. The law has several evidence-based features:

First, the law restricts notifications between 10 PM and 6 AM. Adolescence is characterized by heightened reward sensitivity—particularly in peer contexts—alongside still-developing self-regulation (Spear 2013). Notifications make rewards salient, activating reward-processing regions and reinforcing habitual checking (Sherman et al. 2016). Nighttime social media use is also linked to poorer sleep outcomes, especially among adolescent girls (Alonzo et al. 2021). Limiting nighttime notifications addresses both reward-driven engagement and sleep disruption.

Second, the law allows users to opt into chronological feeds. While algorithmic feeds can support identity exploration by exposing youth to new content, they can also intensify harmful social comparison (de Vries et al. 2016). Chronological feeds give adolescents greater control over content exposure while preserving exploratory benefits. Experimental evidence suggests such feeds can meaningfully reduce use: time spent declined by 20% on Facebook and 10% on Instagram when chronological options were introduced (Guess et al. 2023).

Finally, the law requires youth-accessible reporting mechanisms for harms. This directly addresses a known harm of social media. Studies robustly show that cyberbullying has distinct and additional impacts beyond traditional bullying (Bonanno & Hymel 2013; Nesi et al. 2018). Further, this regulation reflects what youth want. When surveyed about desired resources, the top three youth responses centered around reporting harms: 53% wanted information on blocking people, 52% on reporting mechanisms, and 46% on protecting themselves from uncomfortable experiences (Thorn 2025).

However, good design alone is insufficient. A law's effectiveness ultimately depends on implementation quality. Evidence from Louisiana and China demonstrates that youth and companies often find ways to circumvent restrictions. In practice, Nebraska's legislation may face some of the same pitfalls. To strengthen its approach, Nebraska would benefit from:

1. Including clear compliance monitoring and a transparent auditing process: This may require state-level technical capacity or coordination with federal agencies.
2. Investment into designing age verification technologies and systems: Digital IDs in Louisiana allowed the state to robustly control access to Pornhub. Expansion of programs such as this is outside of the scope of an individual piece of legislation, but could be implemented at a state or federal level.
3. Implementation research funding: Dedicated resources to track outcomes, document circumvention strategies, identify unintended consequences, and capture young people's experiences. Other states could learn from Nebraska's experience only if data are systematically collected.

Next Steps and Policy Implications

Designing comprehensive social media regulations is a challenge, as the Nebraska case demonstrates. However, it is important to remember that regulators have only been addressing this issue for a relatively short time. And often, effective regulatory techniques take time to develop. Consider the example of automobiles: the first car was built in 1886, but seatbelts did not become standard until 1955. Similarly, although social media has been part of our lives for nearly two decades, our understanding of its mechanisms and effects has grown considerably since Facebook first appeared. Current evidence indicates that the impacts of social media are neither uniformly harmful nor universally beneficial. Outcomes are shaped by platform design, context, and individual circumstances.

This analysis contends that the path to more effective regulation lies in shifting from "sledgehammer" approaches—blanket bans, rigid time limits, broad age restrictions—toward

"scalpel" regulations targeting specific design features. While politically expedient, broad restrictions eliminate benefits alongside harms and may disproportionately impact vulnerable youth. Design-based approaches are more difficult to legislate and enforce, but they offer the possibility of reducing documented harms while preserving valuable uses.

Further, effective regulation must be grounded in developmental science. Research on prefrontal cortex maturation, reward sensitivity, and social cognition can inform age-appropriate design standards (Crone & Konijn 2018; Shanmugasundaram et al. 2023). Policymakers must also recognize adolescent diversity. Differences in pubertal timing, identity development, cultural background, and family context shape how youth experience social media (Odgers et al. 2020). Regulations that ignore this diversity risk harm those most in need of protection.

Critically, policymakers must distinguish between theoretical and practical effectiveness. Evidence from Louisiana and China demonstrates that circumvention can undermine well-intentioned laws. Ongoing evaluation, comparative research across jurisdictions, and careful monitoring of unintended consequences are essential. One example of such work is the Neely Social Media Index, which tracked user experiences across platforms over multiple years and was able to document Reddit's evolution from a community space toward an information platform (Xing et al. 2025). Policymakers need similar data for regulatory interventions—not just whether platforms implemented required changes, but whether youth mental health improved, whether harmful experiences decreased, and whether beneficial uses were preserved.

Finally, youth themselves must inform policy development. Adolescents recognize both the risks and benefits of social media and consistently prefer tools that empower them—such as blocking, reporting, and content controls—over blanket restrictions (Badillo-Urquiola et al. 2019; Popat & Tarrant 2023). Policies developed without youth input risk addressing adult anxieties rather than adolescent realities.

Grounding legislative efforts in developmental research, learning from

successes and failures across jurisdictions, ensuring legislation is enforceable, and genuinely centering youth voices in policy development is a tall order. However, these changes are necessary. The stakes are substantial. An entire generation is navigating an unprecedented digital environment during a critical developmental period. We owe younger generations policy responses that meet their reality. “Scalpel approaches,” designed-focused regulations that are targeted, evidence-based, and practical to enforce, offer the most meaningful path forward. The goal is not to restrict youth from the digital world, but to shape the conditions under which they inhabit it. While challenging, it is an achievable standard, and one that policymakers must now commit to meeting.

References

- Alonzo, R., J. Hussain, S. Stranges, and K. K. Anderson. 2021. "Interplay between Social Media Use, Sleep Quality, and Mental Health in Youth: A Systematic Review." *Sleep Medicine Reviews* 56: 101414.
- Avcı, H., L. Baams, and T. Kretschmer. 2025. "A Systematic Review of Social Media Use and Adolescent Identity Development." *Adolescent Research Review* 10 (2): 219–236.
- Badillo-Urquiola, K., D. Smriti, B. McNally, E. Golub, E. Bonsignore, and P. J. Wisniewski. 2019. "Stranger Danger! Social Media App Features Co-Designed with Children to Keep Them Safe Online." In *Proceedings of the 18th ACM International Conference on Interaction Design and Children*, 394–406.
- Berger, M. N., M. Taba, J. L. Marino, M. S. Lim, and S. R. Skinner. 2022. "Social Media Use and Health and Well-Being of Lesbian, Gay, Bisexual, Transgender, and Queer Youth: Systematic Review." *Journal of Medical Internet Research*. 24 (9): e38449.
- Bernard, T. 2024, May 9. "Age Verification in the United States: Insights from the Open Technology Institute." *TechPolicyPress*. <https://www.techpolicy.press/age-verification-in-the-united-states-in-sights-from-the-open-technology-institute/>
- Best, P., R. Manktelow, and B. Taylor. 2014. "Online Communication, Social Media and Adolescent Wellbeing: A Systematic Narrative Review." *Children and Youth Services Review*. 41: 27–36.
- Bonanno, R. A., and S. Hymel. 2013. "Cyber Bullying and Internalizing Difficulties: Above and Beyond the Impact of Traditional Forms of Bullying." *Journal of Youth and Adolescence*. 42 (5): 685–697.
- Brennen, S. B., and Z. Sanderson. 2024. *The State of State Technology Policy: 2024 Report*. NYU Center on Technology Policy and NYU Center for Social Media and Politics. https://techpolicyny.org/wp-content/uploads/2024/12/state-tech-policy_2024_CTP_CSMaP_final.pdf
- California Age-Appropriate Design Code Act, Cal. Civ. Code § 1798.99.28 et seq. (2022). https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202120220AB2273
- Centers for Disease Control and Prevention. 2021. *Youth Risk Behavior Surveillance System: High School Youth — United States, 2021 [Data set]*. Youth Online. <https://youthonline.cdc.gov/#/data?topics=ALL&location1=XX&year1=2021>
- Chen, J., C.-H. Lin, and G. Chen. 2021. "A Cross-Cultural Perspective on the Relationships among Social Media Use, Self-Regulated Learning and Adolescents' Digital Reading Literacy." *Computers & Education*. 175: 104322.
- Crone, E. A., & Konijn, E. A. (2018). *Media Use and Brain Development During Adolescence*. *Nature Communications*, 9(1), 588.
- Daniels, S. R., C.-C. Yang, S. J. Toohey, and V. W. Willard. 2021. "Perspectives on Social Media from Adolescents and Young Adults with Cancer." *Journal of Pediatric Oncology Nursing*. 38 (4): 225–232.
- de Vries, D. A., J. Peter, H. de Graaf, and P. Nikken. 2016. "Adolescents' Social Network Site Use, Peer Appearance-Related Feedback, and Body Dissatisfaction: Testing a Mediation Model." *Journal of Youth and Adolescence*. 45 (1): 211–224. <https://doi.org/10.1007/s10964-015-0266-4>
- Fardouly, J., and L. R. Vartanian. 2016. "Social Media and Body Image Concerns: Current Research and Future Directions." *Current Opinion in Psychology*. 9: 1–5.
- Ferguson, C. J. (2025). *Do social media experiments prove a link with mental health: A methodological and meta-analytic review*. *Psychology of Popular Media*, 14(2), 201.

- Guess, A. M., N. Malhotra, J. Pan, P. Barberá, H. Allcott, T. Brown, ... and J. A. Tucker. 2023. "How Do Social Media Feed Algorithms Affect Attitudes and Behavior in an Election Campaign?" *Science*. 381 (6656): 398–404.
- Hancock, J., Liu, S. X., Luo, M., & Mieczkowski, H. (2022). Psychological well-being and social media use: A meta-analysis of associations between social media use and depression, anxiety, loneliness, eudaimonic, hedonic and social well-being. *Anxiety, Loneliness, Eudaimonic, Hedonic and Social Well-Being* (March 9, 2022).
- Health and Human Services. 2023a. Our Epidemic of Loneliness and Isolation: The U.S. Surgeon General's Advisory on the Healing Effects of Social Connection and Community. U.S. Department of Health and Human Services.
- Health and Human Services. 2023b. Social Media and Youth Mental Health: The U.S. Surgeon General's Advisory. U.S. Department of Health and Human Services.
- Information Commissioner's Office. (2021). Age appropriate design: A code of practice for online services. <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/childrens-information/childrens-code-guidance-and-resources/age-appropriate-design-a-code-of-practice-for-online-services/>
- Issue One. 2024, September 10. "New Poll Finds Near Universal Public Support for Bipartisan Legislation Protecting Kids Online." <https://issueone.org/press/new-poll-finds-near-universal-public-support-for-bipartisan-legislation-protecting-kids-online/>
- Kaye, B. 2025, September 1. "Australian report raises concerns over age-verification software ahead of teen social ban" Reuters. <https://www.reuters.com/world/asia-pacific/australian-report-raises-concerns-over-age-verification-software-ahead-teen-2025-09-01/>
- Keles, B., N. McCrae, and A. Grealish. 2020. "A Systematic Review: The Influence of Social Media on Depression, Anxiety and Psychological Distress in Adolescents." *International Journal of Adolescence and Youth*. 25 (1): 79–93.
- Khalaf, A. M., A. A. Alubied, A. M. Khalaf, A. A. Rifaey, A. Alubied, and A. Rifaey. 2023. "The Impact of Social Media on the Mental Health of Adolescents and Young Adults: A Systematic Review." *Cureus*. 15 (8).
- Lang, D. N., B. Listyg, B. V. Ross, A. V. Musquera, and Z. Sanderson. 2025, March. "Do Age-Verification Bills Change Search Behavior? A Pre-Registered Synthetic Control Multiverse [Pre-print]." OSF. <https://doi.org/10.17605/OSF.IO/VP9Z6>
- Marciano, L., M. Ostroumova, P. J. Schulz, and A. L. Camerini. 2022. "Digital Media Use and Adolescents' Mental Health during the COVID-19 Pandemic: A Systematic Review and Meta-Analysis." *Frontiers in Public Health*. 9: 793868.
- MultiState. n.d. "2025 State Government Trifactas." Accessed October 27, 2025. <https://www.multistate.us/resources/2025-state-government-trifactas>
- National Conference of State Legislatures. 2025, April 8. "Social Media and Children 2025 Legislation." <https://www.ncsl.org/technology-and-communication/social-media-and-children-2025-legislation>
- National Press and Public Administration. 2024, December 16. "The '2024 China Game Industry Report on the Protection of Minors' Was Released." https://www.nppa.gov.cn/xxfb/ywdt/202412/t20241216_877443.html
- Nesi, J., S. Choukas-Bradley, and M. J. Prinstein. 2018. "Transformation of Adolescent Peer Relations in the Social Media Context: Part 2—Application to Peer Group Processes and Future Directions for Research." *Clinical Child and Family Psychology Review*. 21 (3): 295–319.

- Odgers, C. L., S. M. Schueller, and M. Ito. 2020. "Screen Time, Social Media Use, and Adolescent Development." *Annual Review of Developmental Psychology* 2 (1): 485–502.
- Orben, A., and A. K. Przybylski. 2019. "The Association between Adolescent Well-Being and Digital Technology Use." *Nature Human Behaviour* 3 (2): 173–82. <https://doi.org/10.1038/s41562-018-0506-1>.
- Popat, A., and C. Tarrant. 2023. "Exploring Adolescents' Perspectives on Social Media and Mental Health and Well-Being: A Qualitative Literature Review." *Clinical Child Psychology and Psychiatry* 28 (1): 323–37.
- Shannon, H., K. Bush, P. J. Villeneuve, K. G. Hellems, and S. Guimond. 2022. "Problematic Social Media Use in Adolescents and Young Adults: Systematic Review and Meta-Analysis." *JMIR Mental Health*. 9 (4): e33450.
- Shanmugasundaram, M., and A. Tamilarasu. 2023. "The Impact of Digital Technology, Social Media, and Artificial Intelligence on Cognitive Functions: A Review." *Frontiers in Cognition* 2: 1203077.
- Sherman, L. E., A. A. Payton, L. M. Hernandez, P. M. Greenfield, and M. Dapretto. 2016. "The Power of the Like in Adolescence: Effects of Peer Influence on Neural and Behavioral Responses to Social Media." *Psychological Science*. 27 (7): 1027–1035.
- Spear, L. P. 2013. "Adolescent Neurodevelopment." *Journal of Adolescent Health* 52 (2): S7–S13.
- Spies Shapiro, L. A., and G. Margolin. 2014. "Growing Up Wired: Social Networking Sites and Adolescent Psychosocial Development." *Clinical Child and Family Psychology Review*. 17 (1): 1–18.
- Steinberg, L., G. Icenogle, E. P. Shulman, K. Breiner, J. Chein, D. Bacchini, ... and H. M. Takash. 2018. "Around the World, Adolescence Is a Time of Heightened Sensation Seeking and Immature Self-Regulation." *Developmental Science*. 21 (2): e12532.
- Supreme Court of the United States. 2025, June 27. *Free Speech Coalition, Inc. v. Paxton*, 606 U.S. https://www.supremecourt.gov/opinions/24pdf/23-1122_3e04.pdf
- Texas Legislature. 2023. H.B. 1181: An Act Relating to Restricting Access to Sexual Material Harmful to Minors on an Internet Website. (88 R Sess.). <https://capitol.texas.gov/tlodocs/88R/billtext/html/HB01181H.htm>
- Thorn. 2025. *Youth Perspectives on Online Safety, 2024: An Annual Report of Youth Attitudes and Experiences*. https://info.thorn.org/hubfs/Research/Thorn_24_YouthMonitoring_Report.pdf
- Twenge, J. M. 2020. "Increases in Depression, Self-Harm, and Suicide among U.S. Adolescents after 2012 and Links to Technology Use: Possible Mechanisms." *Psychiatric Research and Clinical Practice*. 2 (1): 19–25.
- Valkenburg, P. M. 2022. "Social Media Use and Well-Being: What We Know and What We Need to Know." *Current Opinion in Psychology*. 45: 101294.
- Vilarinho-Pereira, D. R., A. A. Koehler, and D. de Souza Fleith. 2021. "Understanding the Use of Social Media to Foster Student Creativity: A Systematic Literature Review." *Creativity: Theories—Research—Applications*. 8 (1): 124–147.
- Xing, E., Y. Lui, R. Iyer, and N. Fast. 2025, July. "Reddit in Transition: A Two-Year Examination of Social Media Experiences and Evolution." *Psychotech*. <https://psychotech.substack.com/p/reddit-in-transition-a-two-year-examination>
- Zhang, L., and K. Kollnig. 2024. "Theory and Practice: The Protection of Children's Personal Information in China." *International Data Privacy Law*. 14 (1): 37–52. <https://doi.org/10.1093/idpl/ipad017>
- Zhao, Y. 2024, August 20. "Parents Help Children Dodge Time Limits on Online Games." *China Daily*. <https://global.chinadaily.com.cn/a/202408/20/WS66c3e9a1a31060630b923e74.html>

Zhangshao, T., B. Egliston, and M. Carter.
2024, December 10. "China Restricted
Young People from Video Games. But
Kids Are Evading the Bans and Getting
into Trouble." The Conversation.[https://
theconversation.com/china-restricted-
young-people-from-video-games-but-kids-
are-evading-the-bans-and-getting-into-
trouble-245264](https://theconversation.com/china-restricted-young-people-from-video-games-but-kids-are-evading-the-bans-and-getting-into-trouble-245264)

Reciprocal Soft Power Between China and Its Developing Partners

An Examination of South-South Cooperation Rhetoric

Odessa Ikels¹

Abstract

There has been increasing global awareness of China's efforts to expand its soft power and economic statecraft in the Global South. Chinese policy has become spotlighted by critics in the United States and Europe, who often use the label "debt-trap diplomacy" to highlight the predatory nature of China's infrastructure grants, mining projects, and other investments. This research asks the question of why, if news of China's "traps" is becoming widespread, leaders of Global South nations continue to reach out to China and agree to deals with Chinese state-owned enterprises (SOEs). An interesting case in the chain of Chinese investment projects is the construction of a new national stadium and national library in El Salvador. Analysis of the implementation and state discourse surrounding this project demonstrates key mutual benefits for aspiring regional powers. This research challenges the erasure of Global South agency and attempts to offer a clearer picture of the motivations of nations like El Salvador in forming a relationship with China. The dual, but unequal, benefits conferred by partnerships between China and its Global South partners are central to understanding the success of its Belt and Road Initiative (BRI) and Chinese soft power more broadly.

China's rhetoric of South-South Cooperation (SSC) is often dismissed as window-dressing in its relationships with the developing world. In this study, I analyze the rhetoric and argue that it should be taken seriously for two reasons. First, as the language of public diplomacy, rhetoric can shape how people think of China and its relationships. Second, it reveals possible places of genuine "mutual benefit" that can explain why developing countries seek out and maintain their relationships to China. I make this argument by examining China's relationship with El Salvador and its SSC infrastructure projects, which can be understood as "symbols of state" by international relations scholar Austin Strange's definition. He argues that countries with little regional influence and recent changes in leadership seek out these prestige projects with China to boost their domestic and regional legitimacy, through their association with China (Strange 2024). These projects benefit their government and elite class, but do not materially help most of the populace. Studying El Salvador's prestige projects allows us to identify how China and the Salvadoran government have complimentary goals of increasing their influence in El Salvador and in Central America more broadly. While the beginning of their diplomatic linkage can be understood as primarily economic statecraft,

their current and future relationship should be understood as driven by their mutually beneficial soft power goals.

South-South Cooperation Scholarship: Superficial or Genuine?

In debates about China's South-South Cooperation rhetoric and policies, scholars focus on China's discourse of "mutual benefit," "shared interests," and broader messaging about the shared experience of colonization of fellow developing countries. China uses this language to dress up their infrastructure projects and investment proposals to Global South countries, and the rhetoric is often co-produced by the developing countries themselves. Scholars often dismiss China's rhetoric to focus on the effects of its development projects, which enable China to access natural resources, new markets, and to increase international diplomatic support of the People's Republic of China (PRC) over the Republic of China (ROC). While China does pursue these goals, some researchers have begun to question whether previous academic approaches have adequately accounted for the agency of Global South countries in seeking out projects with China, and how infrastructure projects can create a positive national image of

¹ Odessa Ikels is a student at Amherst College

China in host countries, at least in the short term. Studying prestige infrastructure projects as soft power instruments of both China and its target nations can yield a more comprehensive understanding of partnerships under the South-South Cooperation banner.

In their study of South-South Cooperation as rhetoric, many scholars see SSC as superficially employed by China in its diplomatic relationships with the Global South, intended to create a shared discourse that ameliorates possible partners in its quest for natural resources and international rejection of Taiwan's sovereignty. This economic statecraft approach is characterized by skepticism that China is meaningfully changing the international system, since China seeks to create new economic zones for itself and access to markets, leaving the neoliberal system fundamentally unchallenged (Gülseven 2023; Mohan 2016; Gonzalez-Vicente 2017; Mendes 2024; Zhang 2020). These scholars critique the language of SSC for its "open-ended" conceptual basis and its appeals to solidarity based on a shared identity of colonization and exploitation, ultimately enabling China's exploitation of these countries (Mohan 2016). Gonzalez-Vicente similarly sees China's SSC as reproducing processes of exploitation, but highlights the elite vs. non-elite relationship, where everyday people are often on the losing end of resource deals with China (Gonzalez-Vicente 2017). Mendes' study builds on Gonzalez-Vicente's, arguing that China's SSC rhetoric was effective at bringing African political elites into agreements that would benefit them (Mendes 2024). These scholars all acknowledge how soft power in technology, health aid, and other areas are interlinked with SSC as part of broader Chinese foreign policy, but their approach fundamentally treats SSC as the rhetorical exterior of a materialistic economic statecraft campaign by China, rather than a discourse with intrinsic power or influence.

While a more instrumentalist lens offers key insights about the ways that Chinese investments often involve some political alignment with China, a soft power approach can help explain why countries continually reach out to China for projects and align themselves with China beyond the initial diplomatic deals. Joseph Nye, who coined the term, originally defined "soft power" as a type of power "which occurs when

one country gets other countries to want what it wants" (Nye 1990, 166). Hayden builds on Nye's conceptualization of soft power, arguing that the rhetoric and public diplomacy of a country's officials are key tools of soft power, which can shape national image to foreign publics (Hayden 2011). Repnikova adapts Nye's concept of soft power to the Chinese context, arguing that Chinese soft power includes "pragmatic enticements," and articulates how Chinese conceptualizations of soft power are much more fluid, broadly encompassing an emphasis on "culture, moral principles, and political ideology" (Repnikova 2022, 4). Soft power in the SSC context should be understood as possibly functioning on two levels. First, that the language of SSC both reveals key complexities of the relationship between China and its developing partners, and that the rhetoric influences China's image in the host country and has the potential for subject formation. Second, the literal infrastructure projects described by SSC language can create positive feelings towards China and Chinese approaches to development in host countries, which could create or reinforce a positive relationship with China. It is also helpful to recognize that the SSC rhetoric, and Chinese soft power in general, are situated in the context of Beijing's desire to restructure the world order through multilateral organizations, like BRICS and the Association of Southeast Asian Nations (ASEAN), as well as infrastructure deals through the BRI, the Regional Comprehensive Economic Partnership (RCEP), and the Asian Infrastructure Investment Bank (AIIB). Regional soft power in Latin America, therefore, must be understood as contributing to China's larger geopolitical strategy.

A number of scholars take a constructivist approach to China's identity creation through its SSC narratives, where SSC demonstrates China's simultaneous effort to create an identity as both a fellow Global South country and a major power with resources to fund development (DeHart 2012; Bocchese and Linn 2016; Hönke and Yang 2024; Zhang 2020; Mann 2023). DeHart argues that Chinese policymakers create a dual, conflicted identity through SSC where "China appears simultaneously as a First World donor and the quintessential Third World labourer" (DeHart 2012, 1371). Soft power then becomes a way to stabilize China's identity as a Global

South country and the SSC rhetoric is therefore a condition of possibility for China to pursue its economic and political goals (Bocchese and Linn 2016). This identity creation, instead of being flimsy procedural language in China's agreements, substantively allows it to participate in relations with developing countries and is a precondition to engage in infrastructure deals. Many of these authors agree that in specific contexts, Chinese projects under the SSC banner initially have positive receptions, but host countries' perceptions sour after implementation (DeHart 2012; Bocchese and Linn 2016; McCauley et al. 2022). Comprehensive studies have yet to identify whether the vague nature of SSC language, which constructs a strategically ambiguous identity for the Chinese government and its SOEs, plays a role in creating opposition to the Chinese model of development over time (Hönke and Yang 2024).

Many scholars study the agency of developing countries in their relationships with China, which exposes the inability of an economic statecraft approach to explain why those countries would initiate relationships and non-traditional infrastructure projects under SSC. Scholars and diplomats in the West posit that developing countries often receive poor infrastructure that does not meet their development needs or investments that only benefit China (Mendes 2024; DeHart 2012), which the economic statecraft approach often accepts at face value. Strange demonstrates that in some cases, the developing country with pressing infrastructure needs paradoxically reaches out to China to cooperate over prestige projects, such as stadiums and highly visible buildings. Through association with China and the construction of these buildings, the developing country is able to increase their domestic and regional legitimacy. Therefore, both China and the government of the developing country gain a form of soft power through this form of SSC project (Strange 2024). More broadly, this agency-oriented group of scholars reveal that governmental elites within developing countries benefit from relationships with China by gaining their own soft power and legitimacy from these prestige projects (Bocchese and Linn 2016; Gonzalez-Vicente 2017; Strange 2024). How the soft power and legitimacy interests of developing countries shape the SSC dynamic remains largely

unexplored, begging the question of whether SSC rhetoric can reveal some degree of mutual benefit for the government elites in developing countries and China.

Finally, some scholars argue that Chinese SSC rhetoric ought to be understood as signaling relationality and reciprocity, where China seeks to create a long-term, soft power relationship through a cycle of exchange that goes beyond a bounded, instrumental deal. Scholars emphasize that Chinese relational theory, and *guanxi*, is key to understanding how, from the Chinese perspective, linkages between countries form relationships beyond a simple exchange of economic support for political alignment and actually construct shared identities and interests (Hönke and Yang 2024; Rudyak 2023). One of the central gaps in the literature, however, is understanding how cooperating countries and China understand these relationships differently. If a cycle of gift giving is key to maintaining these relationships, why do we see declining attitudes towards China after the infrastructure has been constructed (Rudyak 2023; McCauley et al. 2022)? More research could focus on whether negative perceptions after implementation bracketed out China permanently or were more of a short-term blip that China could correct. Can relationships between China and developing partners persist beyond the initial moment of the infrastructure gift? Must there be soft power benefits for elites with influence in both countries for these relationships to continue?

While it is imperative to understand how China's economic incentives factor into its deployment of SSC rhetoric, examining how the discourse itself reveals complementary benefits in its relationships is key to understanding China's relationship to the Global South, as both a self-proclaimed member and lender. Research on the question of agency and dual soft power that China and target countries seek out through SSC presents a particularly interesting research direction to fully understand how relationships between China and developing nations evolve beyond initial investments.

Prestige Projects and the Promise of Mutual Benefit

In the last twenty years, China has increased its investment in stadiums and highly visible buildings in the Global South. These “prestige projects” offer a lens through which we may determine whether China’s promises of “mutual benefit” to the developing world match its material implementations (Strange 2024). El Salvador’s cooperation with China, given their recent diplomatic alignment, illustrates how Global South countries co-produce the rhetoric of SSC and “friendship” alongside China. China’s recent investment in prestige projects in El Salvador demonstrates how China’s soft power ventures enhance its own interests and serve the interests of the target country’s government. Granted, the interests of the governments of developing countries do not always represent those of its people, but the mutual benefits conferred by these relationships can explain how and why developing countries engage. This section pushes back on narratives of top-down Chinese influence in the developing world, demonstrating how the rhetoric of SSC reveals complementary soft power objectives that define the current China-El Salvador relationship. To do so, I provide an overview of the political context before discussing two levels of soft power: first I examine the soft power exerted by the projects themselves, then I argue that the rhetoric is inherently influential and reflects complexities in the relationship. Finally, I discuss the broader scholarly implications of a soft power approach to South-South Cooperation relationships.

Political Context in El Salvador

El Salvador presents a compelling case study to examine China’s prestige projects and its relationship with its developing partners. El Salvador itself is relatively small, and has not had great amounts of influence in Central America historically, only recently attracting international attention because of President Nayib Bukele’s changes to the political system and contentious foreign policy with respect to the United States and China (Ellis 2021). He is domestically bolstered by high approval ratings and growing popularity in both the Central American region and globally for his

strong crackdown on gangs and violent crime (Bergengruen 2024). While Bukele initially campaigned with strong anti-China messaging in 2019, he has since pivoted to deepening the previous Ceren administration’s diplomatic connections with the Chinese government (Duckworth 2022). Bukele’s increased international status and visibility in the last six years suggests that we should closely examine its relationship with China to determine the role played initially by the prestige projects in the country to give it increased status. Additionally, Central America was once a region with strong support for the Republic of China (ROC), which the PRC has progressively eroded. El Salvador was particularly strategic for China given its role as a rising regional leader: Nicaragua, Honduras, the Dominican Republic, and other countries followed El Salvador’s 2018 renunciation of Taiwan and recognition of the PRC (Roy 2025).

El Salvador’s rapidly evolving relationship with the People’s Republic of China is in fact intrinsically linked to its diplomatic rejection of Taiwan, which the country had previously recognized for over 80 years, in 2018 under the Ceren administration (Wang 2018). The Ceren administration cut ties with Taiwan by recognizing the PRC and its “One China Policy,” which declares that the ROC is part of the PRC’s inalienable territory (Lemus-Delgado and Dueñas 2021). After this recognition, China publicly promised USD 150 million in development investments in El Salvador, signaling a clear *quid pro quo*, although no projects were signed (Lemus-Delgado and Dueñas 2021).

Under Bukele, the Chinese government formally agreed to build and fund an additional USD 500 million worth of investment projects in El Salvador, including a new national stadium and library, updates to tourist attractions, and other projects—notably infrastructure that is highly visible rather than crucial to development needs (Duckworth 2022). While the investment did include a water facilities project, it has been noticeably sidelined in the press and is not Bukele’s priority (Ellis 2021). The continued cooperation with China has frustrated officials in Washington, who repeatedly met with the Ceren and Bukele administrations to try to convince El Salvador to reestablish ties with Taiwan (Lemus-Delgado and Dueñas 2021, 63). While Bukele’s reclusiveness makes his intentions and

motivations unclear, his actions, which build on Ceren's diplomatic advancements, suggest that a long-term relationship with China is in the interests of El Salvador, beyond a single administration's goals.

Prestige Projects as Soft Power Tools for Both Governments

Chinese infrastructure projects in El Salvador can be understood as "prestige projects," according to Strange's framework, and we can examine their relevance to China's soft power in the region in that context. The investments in a new national stadium and library for El Salvador, are consistent with Strange's articulation of prestige projects, which he defines as "international development projects with high visibility and national symbolism" (Strange 2024, 3). The new National Library of El Salvador, abbreviated as BINAES, opened to the public in 2023 and flies the Chinese flag outside its entrance. Its pyramidal design, however, is intended to reflect Salvadoran heritage through its pre-Columbian architecture (Virginia 2024). Adding to its level of prestige and visibility, BINAES is also the most modern library in Latin America and largest in Central America (Bukele 2023). Similarly, Bukele has proclaimed that the new national stadium, which is currently under construction and set to open in 2027, will be the largest in Central America, with capacity of over 50,000 seats, and also the most modern in Latin America as a whole (Bukele 2023). Bukele has worked to tie these highly visible projects to the national image of El Salvador by emphasizing their homage to the Salvadoran past and vision for the future (El Salvador News 2023).

These prestige projects are a tactic of Chinese soft power, as seen in the Costa Rican case studied by Strange, by tying perception of China to modern development and popular buildings that enhance quality of life in the host country (Strange 2024). These projects seem to be relatively unique to China, as the United States and other international bodies typically only fund projects necessary for direct economic development. The use of private funding is also distinct as its ties to soft power are more attenuated. The connection between the new infrastructure and China is explicit in the case of the BINAES, which blatantly ties the image of China, through its flag, to a symbol of Salvadoran modernity, education,

and national architectural heritage. The stadium has yet to open, and its impact is therefore hard to measure, but the addresses delivered by President Bukele and the Chinese Ambassador to El Salvador, Zhang Yanhui, indicate that it will have similar visual messaging (Bukele 2023). Additionally, during the ceremony of laying the first cornerstone of the stadium, Zhang and Bukele gave speeches standing side by side in front of the worksite, with the Chinese and Salvadoran flags behind them, the image framing China as a bringer of modernity and culture, while being on equal footing with El Salvador (Bukele 2023). The ceremony received hundreds of thousands of views on Twitter alone (Bukele 2023), demonstrating the impact of such visual messaging on connecting the image of China with positive developments in Salvadoran quality of life.

The Salvadoran government stands to gain legitimacy and influence, both regionally and domestically, from its relationship with China, according to Strange's prestige project framework. The prestige projects in El Salvador are typical of Strange's framework; they are grandiose, aim to improve the national image of both China and El Salvador, and legitimize and create regional influence for Bukele's government. According to analysts, El Salvador has the capacity to influence other states in Central America based on how it juggles the interests of the PRC and the United States in the region (Ellis 2021; Lemus-Delgado and Dueñas 2021). Bukelismo, or the Bukele approach and posture, has become increasingly attractive in Latin America based on his aforementioned crackdown on crime and has earned him a growing presence in global media (Bergengruen 2024). Simultaneously, scholars emphasize how Bukele's high approval ratings are at least partially a product of his political intimidation, and he receives negative press from global news sources for his authoritarian tendencies (Gellman 2022). His involvement in Chinese-backed projects, which are the largest of their kind in Central America and most modern in Latin America (Bukele 2023), influences the region to adopt similar policies and align themselves with him to reap similar benefits and gain his support. Bukele's position as a recently elected leader, who is changing El Salvador's domestic and foreign policy, means that he stands to gain more press, legitimacy, and power

in the region from Chinese-backed prestige projects.

The Soft Power of “South-South Cooperation” Rhetoric

Not only are the physical buildings and ceremonies attempts at increasing Chinese soft power in El Salvador, but the South-South Cooperation rhetoric used to describe these projects is also powerful in itself. In the brick-laying ceremony, Bukele and Ambassador Zhang emphasized that the stadium will be a symbol of the “friendly cooperation” between the countries (Bukele 2023). In most official documentation and communication regarding the relationship between El Salvador and China, Chinese diplomats emphasize this idea of “friendship” between the two countries (Xinhua 2018a; BBC 2019; Bukele 2023; Xinhua 2018b). This discourse of “friendship” and “mutual benefit” exerts soft power to shape the subject formation of Salvadorans, attempting to positively impact how they perceive China. In a joint communique between the Bukele and Chinese governments in 2019, the countries declared that they would approach the relationship with “mutual respect, equal treatment, cooperation for win-win outcome, and non-interference of each other’s internal affairs,” and would “continuously enhance political mutual trust, conduct friendly exchanges, [and] strengthen pragmatic cooperation” (BBC 2019). This rhetoric is consistent with China’s broad SSC discourse used in the context of its relationship to the developing world, which emphasizes the mutual benefits of cooperation. Scholars often argue that this rhetoric is transparently superficial, and that claims of an equal relationship are weak given the difference in political power, economic might, and outcomes of previous relationships between China and developing countries (Gülseven 2023). However, in this joint statement, the countries state that relations will be “an example of friendly cooperation between countries of different scale and national conditions” and will “promote the deepening of South-South cooperation” (BBC 2019). In fact, the governments explicitly acknowledge the differences between the nations but insist that friendship and “win-win outcome[s]” can still be achieved, even if the wins are proportional according to the preexisting economic “conditions” (BBC 2019). This is not

utopian rhetoric, but a real acknowledgement of their respective situations and a promise of proportional benefit. This tactful recognition of the power imbalance in the relationship might lead us to conclude that SSC rhetoric not only exerts soft power in itself but also contains elements of truth about the El Salvador-China relationship as effective for the Salvadoran government’s regional aspirations, which merits greater scholarly consideration than is currently given.

Broadly, Bukele’s rhetoric describing El Salvador’s relationship with China, which dovetails with China’s own language, demonstrates how he aims to gain soft power and legitimacy in El Salvador and regionally through the diplomatic linkage. In one of his earliest speeches about the relationship to China in 2019, Bukele stressed that he was “committed to developing a long-term friendship with China and developing relations with China in parallel with other major bilateral relations, so as to set an example for other countries in Latin America,” demonstrating the co-authored nature of the SSC rhetoric (Embassy of the PRC 2019). Bukele aims to be a regional leader and creates status for himself through diplomatic relations with China and also by using language that positions himself as “an example” of someone worth emulating (Embassy of the PRC 2019). In his address at the brick-laying ceremony, he interwove references to his achievements regarding China with his successes at taking El Salvador from the most dangerous country in the world, to one of the safest in the region (Bukele 2023). By linking his growing regional popularity from Bukelismo to his collaboration with China, he reinforces his legitimacy and frames his government as successful at both solving domestic problems and making powerful friends internationally. The rhetoric describing the prestige projects, therefore, is a form of soft power in itself, intended to increase Bukele’s regional influence as well as China’s. Additionally, Bukele mitigates the image of himself as a dangerous, unpredictable dictator by emphasizing the reliable, “long-term” nature of the El Salvador-China “friendship” he has fostered and pointing to concrete benefits of his restrictive policies (Bukele 2023; Embassy of the PRC 2019; BBC 2019). This language attempts to make his administration seem worth partnering with or mimicking. The co-created SSC rhetoric,

borne out of these joint statements and Bukele's independent mirroring of China's rhetoric, is therefore intended to express the agency and benefit to both countries. The language reflects that Bukele stands to gain much from the relationship with China in the form of increased influence in the region as well as from the direct benefits of the projects themselves.

Reception of the relationship between China and El Salvador, while limited given the media conditions under Bukele's regime, has been positive and indicates that both China and Bukele's government benefit from the projects. According to some non-profit organizations, the projects have positively impacted public opinion, portraying Bukele as efficient and China as a bringer of modernity and economic prosperity (Piña 2024). Positive reception of the relationship with China can also be traced by using the Latinobarómetro 2023 data, which determined that as the overall opinion rating of China in El Salvador went up during Bukele's presidency (Piña 2024). Additionally, social media likes, in proportion to national population, further indicate positive reception (Piña 2024; El Salvador News 2023; Bukele 2023). The highly visible nature of these prestige projects transcend traditional and social media: in almost a year and a half since its opening, 2.5 million people have visited the BINAES, which, given El Salvador's population of 6.3 million, confirms the influence of the building. The library hosted Chinese New Year celebrations in 2024 (Fátima Membreño and René Gómez 2024), promoting Chinese culture and Salvadoran modernity in tandem. Unfortunately, while we can attempt to triangulate the reception of the soft power projects, the fact that few Salvadoran opinions are published in the media makes tracing the impact, at this point in time, extremely difficult.

Soft Power in the Broader Context of China-Developing World Relations

By examining their joint statement from 2019, we can identify the dual nature of the China-El Salvador relationship as simultaneously containing components of economic statecraft and soft power goals. The most explicit instrumental component of the relationship was the initial rejection of recognition for Taiwan by El Salvador, which was followed by China's investments (Bergengruen

2024). The joint statement released by Xinhua, which mostly uses SSC rhetoric to reinforce the idea of friendship between the two nations, also contains a few paragraphs which seem to be most explicitly forming a quid pro quo between the countries. In exchange for China's development investments, El Salvador promised to help expand Chinese influence in the region:

The El Salvador side will participate actively in the building of the China-CELAC (Community of Latin American and Caribbean States) Forum and promote general cooperation between China and Latin America, so as to contribute its efforts toward the China-Latin America comprehensive cooperative partnership of equality, mutual benefit, and common development to bear new fruits in the new era. (BBC 2019)

The explicit demand to get access to other Latin American countries through CELAC is certainly an instrumental component of the relationship. However, the ambiguous nature and difficulty assessing the demands of "friendship," as well as the relatively low Salvadoran economic dependence on China, suggest that the relationship is less straightforward than it might appear. From an economic statecraft perspective, for China to exert long-term political influence over El Salvador, it needs tangible leverage in the Salvadoran economy. However, unlike many of China's other relationships in Latin America, El Salvador does not export a crucial resource to China, its main exports being coffee and sugar (Ellis 2021). The new diplomatic relationship, as of 2021, has resulted in a greater Chinese presence in El Salvador's economy in the form of cheap imports (Ellis 2021). However, these developments are more directly attributable to the effects of greater economic linkages rather than a demonstrably dependent relationship, with China only being El Salvador's ninth biggest export destination even after trade increased as a result of the 2019 agreement (Colantuoni 2024; Prensa Latina 2024). Though continuing negotiations for a free trade agreement may have been catalyzed by the initial linkage, the fact that it is forthcoming indicates that factors other than pure economic leverage motivate the partnership between the countries. While the actions taken to initiate the relationship—an exchange between recognizing the PRC and gaining development infrastructure, with both sides gaining regional influence from

the relationship—could be considered merely transactional, the countries' future relationship remains largely up in the air. China simply does not have as much leverage in El Salvador as it does in other developing countries, and therefore it would be harder to force actions by El Salvador should Bukele attempt to step back from the relationship. Additionally, Bukele has shown that he is unafraid to play the United States and China off of each other to gain the best deal for El Salvador given their competing interests in the region, which adds to the strategic uncertainty (Lemus-Delgado and Dueñas 2021). It is therefore more accurate to understand the future of their relationships as being determined by mutual goals for regional influence rather than dictated by Chinese economic leverage.

The China-El Salvador relationship calls into question how soft power functions within relationships between China and developing countries who seek to increase their influence regionally. The idea that a domestic project in El Salvador would be able to shape the trajectories of surrounding countries seems to suggest that soft power might be more fluid and transferable than scholars have previously imagined. Empirically, infrastructure investments have strengthened favorable perceptions of China within the target nation, but these improved perceptions of the donor country have often been short lived (McCauley et al. 2022). As a result, the impacts of prestige projects are traceable as short-term boosts to popularity with unclear long-term consequences. In the Salvadoran case, Bukele's desire to add to his public image early on in his presidency could reflect his desire to back up his increasingly authoritarian governance with domestic popularity and support from neighboring countries. We can also see how the line between soft power projected regionally and domestic propaganda blurs in Bukele's case. Prestige projects as soft power tools, therefore, might be more advantageous for nascent authoritarian governments or countries with recent major changes to their political system. Future research could explore the possible correlation between countries that have undergone massive political change and countries that seek out prestige projects. While the systemic motivations for engaging with China remain difficult to delineate, this analysis has sought to demonstrate that in the

case of El Salvador, the relationship morphed from being predominantly characterized by economic statecraft to being focused on mutually beneficial regional soft power motivated by long-term strategy (BBC 2019). This transition demonstrates the intertwined nature of economic statecraft and reciprocally constitutive soft power, which reinforce each other and should be used in tandem to describe China's partnerships with developing countries.

A Long-Term Friendship?

While it is hard to pinpoint exactly what the "long-term friendship" will look like between these nations, their rhetoric surrounding the prestige projects indicates that their interests are aligned in their desire to increase their respective prestige in El Salvador and their soft power in the region. While China uses the prestige projects and its rhetoric to increase its soft power in El Salvador, the Salvadoran government is not without agency or benefit in the partnership, as it seeks its own regional influence. This paper seeks to blur the boundary between soft power and economic statecraft as frames of analysis, to show that projects can be tools of soft power for both governments, containing elements of economic statecraft while creating more complex relationships. The governments' rhetoric describing their cooperation as "mutually beneficial" and "friend[ly]" exerts power of its own to shape perceptions of the countries while revealing the complexities of the relationship. While tracing the efficacy of this language and its effect on the populace of El Salvador is challenging given that the national stadium is yet to be completed and El Salvador limits the free expression of its citizens, we can understand how this language exerts power of its own to frame how people see the diplomatic relationship. Future scholarship should therefore be vigilant of instances where countries might be mutually supporting one another beyond their individual interests to examine with greater precision what contemporary "friendship" between states truly means.

References

- BBC. 2019. "Xinhua Releases Full Text of Joint Statement between China, El Salvador." December 12.
- Bergengruen, Vera. 2024. "Nayib Bukele's Iron Fist Has Transformed El Salvador." *Time Magazine*, August 29. <https://time.com/7015598/nayib-bukeles-iron-fist-el-salvador/>.
- Bocchese, Marco, and Elizabeth Linn. 2016. "Pricing Identity: Trade, Soft Power and Chinese Development in the Global South." *SSRN Electronic Journal*, ahead of print. <https://doi.org/10.2139/ssrn.2735889>.
- Bukele, Nayib, dir. 2023. Twitter Broadcast of Ceremony to Lay the First Brick to the National Stadium. Twitter. https://x.com/nayibbukele/status/1730572586860323304?ref_src=twsrc%5Etfw%7Ctwcamp%5Etwembed%7Ctwterm%5E1730572586860323304%7Ctwgr%5E20e0d62f13c22ac095ea1f6f32fd3008f4fa0817%7Ctwcon%5Es1_c10&ref_url=https%3A%2F%2Felsalvadorinenglish.com%2F2023%2F12%2F01%2Fnayib-bukele-lays-foundation-stone-for-el-salvadors-national-stadium%2F
- Colantuoni, Steve. 2024. "China and El Salvador Advance Free Trade Negotiations." *The Central American Group*, November 4. <https://www.thecentralamerican-group.com/china-and-el-salvador-advance-free-trade-negotiations/>.
- DeHart, Monica. 2012. "Remodeling the Global Development Landscape: The China Model and South–South Cooperation in Latin America." *Third World Quarterly* 33 (7): 1359–75. <https://doi.org/10.1080/01436597.2012.691835>.
- Duckworth, Elaine. 2022. *Latin American and Caribbean Participation in China's Belt and Road Initiative*.
- El Salvador News. 2023. "Nayib Bukele Lays Foundation Stone for El Salvador's National Stadium." December 1. <https://elsalvadorinenglish.com/2023/12/01/nayib-bukele-lays-foundation-stone-for-el-salvadors-national-stadium/>.
- Ellis, Evan. 2021. "China and El Salvador: An Update." *Center for Strategic and International Studies*, March.
- Embassy of the People's Republic of China in Myanmar. 2019. "Xi Jinping Holds Talks with President Nayib Armando Bukele Ortez of El Salvador." December 3. http://mm.china-embassy.gov.cn/eng/zgxw/201912/t20191205_1397780.htm.
- Fátima Membreño and René Gómez. 2024. "El Año Nuevo Chino Brilló En El Centro Histórico de San Salvador." *Diario El Salvador*, February 15. https://diarioelsalvador.com/el-ano-nuevo-chino-brillo-en-el-centro-historico-de-san-salvador/463531/#-google_vignette.
- Gellman, Mneesa. 2022. "The Democracy Crisis in El Salvador: An Overview (2019-2022)." *Columbia CeMeCA's Regional Expert Paper Series*, no. 4 (September).
- Gonzalez-Vicente, Ruben. 2017. "South–South Relations under World Market Capitalism: The State and the Elusive Promise of National Development in the China–Ecuador Resource–Development Nexus." *Review of International Political Economy* 24 (5): 881–903. <https://doi.org/10.1080/09692290.2017.1357646>.
- Gülseven, Yahya. 2023. "China's Belt and Road Initiative and South-South Cooperation." *Journal of Balkan and Near Eastern Studies* 25 (1): 102–17. <https://doi.org/10.1080/19448953.2022.2129321>.
- Hayden, Craig. 2011. "Introduction." In *The Rhetoric of Soft Power: Public Diplomacy in Global Contexts*. Lexington Books/Fortress Academic.

- Hönke, Jana, and Yifan Yang. 2024. "Liminally Positioned In/Towards the South: Chinese Relations with Africa." In *Asia-Africa- Multifaceted Engagement in the Contemporary World*, edited by Ute Fendler and Yongkyu Chang. Africa's Global Engagement: Perspectives from Emerging Countries. Springer Nature Singapore. https://doi.org/10.1007/978-981-97-0696-9_6.
- Lemus-Delgado, Daniel, and Carlos Cerda Dueñas. 2021. "China or Taiwan? The Tricky Dilemma of Diplomatic Recognition: The Salvadorean Experience." *Contemporary Chinese Political Economy and Strategic Relations* 7 (1): 45–78.
- Mann, Anant Singh. 2023. "Deconstructing the Metanarratives of China's Engagement with the African Continent: African Agency and Competing Narratives." *India Quarterly: A Journal of International Affairs* 79 (3): 318–35. <https://doi.org/10.1177/09749284231184619>.
- McCauley, John F., Margaret M. Pearson, and Xiaonan Wang. 2022. "Does Chinese FDI in Africa Inspire Support for a China Model of Development?" *World Development* 150 (February): 105738. <https://doi.org/10.1016/j.worlddev.2021.105738>.
- Mendes, Carmen Amado. 2024. "The Implications of China's South–South Cooperation Diplomacy: Consistent or Shifting?" In *The Palgrave Handbook on China-Europe-Africa Relations*, edited by Yichao Li, Francisco José B. S. Leandro, Jorge Tavares Da Silva, and Carlos Rodrigues. Springer Nature Singapore. https://doi.org/10.1007/978-981-97-5640-7_44.
- Mohan, Giles. 2016. "Emerging Powers in International Development: Questioning South–South Cooperation." In *The Palgrave Handbook of International Development*, edited by Jean Grugel and Daniel Hammett. Palgrave Macmillan UK. https://doi.org/10.1057/978-1-137-42724-3_16.
- Nye, Joseph S. 1990. "Soft Power." *Foreign Policy*, no. 80: 153–71. JSTOR. <https://doi.org/10.2307/1148580>.
- Piña, Carlos. 2024. "Ties with China Help Bukele Maintain His Advantage in El Salvador's Upcoming Elections." *La Fundación Andrés Bello – Centro de Investigación Chino Latinoamericano*, February 4. <https://fundacionandresbello.org/noticias/el-salvador-%F0%9F%87%B8%F0%9F%87%BB/los-lazos-con-china-ayudan-a-bukele-a-mantener-su-ventaja-en-las-venideras-elecciones-de-el-salvador/>.
- Prensa Latina. 2024. "El Salvador Increased Exports to China Even without an FTA." September 12. <https://www.plenglish.com/news/2024/09/12/el-salvador-increased-exports-to-china-even-without-an-fta/>.
- Repnikova, Maria. 2022. *Chinese Soft Power. Elements in Global China*. Cambridge University Press. Cambridge Core. <https://doi.org/10.1017/9781108874700>.
- Roy, Diana. 2025. "China's Growing Influence in Latin America." *Council on Foreign Relations Backgrounder* (January). <https://www.cfr.org/backgrounder/china-influence-latin-america-argentina-brazil-venezuela-security-energy-bri>.
- Rudyak, Marina. 2023. "'We Help Them, and They Help Us': Reciprocity and Relationality in Chinese Aid to Africa." *Journal of International Development* 35 (4): 583–99. <https://doi.org/10.1002/jid.3699>.
- Strange, Austin. 2024. "Symbols of State: Explaining Prestige Projects in the Global South." *International Studies Quarterly* 68 (2): 1–15. <https://doi.org/10.1093/isq/sqa049>.
- Virginia, Anne. 2024. "El Salvador UnVeils Ultra Modern National Library Gifted by China." *Le Monde International*, February 9. https://www.lemonde.fr/en/m-le-mag/article/2024/02/09/el-salvador-unveils-ultra-modern-national-library-gifted-by-china_6507335_117.html.

- Wang, Zhicheng. 2018. "El Salvador Chooses China over Taiwan." Asia News, August 21. <https://www.asianews.it/news-en/El-Salvador-chooses-China-over-Taiwan-44707.html>.
- Xinhua. 2018a. "People's Republic of China, Republic of El Salvador Establish Diplomatic Ties." August 21. http://english.scio.gov.cn/topnews/2018-08/21/content_59571933.htm.
- Xinhua. 2018b. "Xi Holds Talks with El Salvador President, Urging Solid Basis to Boost Cooperation." November 1. http://www.xinhuanet.com/english/2018-11/01/c_137575187.htm.
- Zhang, Denghua. 2020. "Chinese Foreign Aid and Financing: An Example of New Development Assistance?" In *New Development Assistance*, edited by Yijia Jing, Alvaro Mendez, and Yu Zheng. Governing China in the 21st Century. Springer Singapore. https://doi.org/10.1007/978-981-13-7232-2_10.

THIS PAGE LEFT INTENTIONALLY BLANK.

Talking With the Mic Off

The Passage of Custodial Interrogation Recording Laws

William Lowthert¹ and Christopher J. Normile²

Introduction

Wrongful convictions remain a persistent feature of the American criminal justice system. Although estimates vary widely, most empirical research suggests that between 1% and 3% of criminal convictions are of innocent individuals, with some estimates placing the figure even higher (Zalman, Rubino, and Smith 2017; Zalman, Smith, and Kiger 2008). When considered in isolation, these percentages may appear modest or unimportant. Yet when applied to the scale of incarceration in the United States, they translate into thousands of individuals wrongfully deprived of liberty each year. In any given year, the Federal Bureau of Prisons incarcerates approximately 450,000 individuals, implying that even the most conservative estimates correspond to several thousand wrongful imprisonments annually (Bureau of Justice Statistics 2023). Such errors pose not only individual harms, but also a systemic challenge to the legitimacy and moral authority of the criminal justice system (Bureau of Justice Statistics 2023). Despite broad expert consensus that recording custodial interrogations can reduce wrongful convictions (Kassin et al. 2014), the adoption of these laws remains uneven across US states. This paper argues that long-term Republican legislative control, especially in racially diverse states, is the primary political factor preventing their passage.

A substantial body of research identifies false confessions during police interrogations as a major contributor to wrongful convictions. Legal scholars and psychologists have repeatedly shown that commonly used interrogation practices, particularly those involving prolonged isolation, deception about evidence, and implicit promises of leniency, can induce confessions from innocent suspects, especially among juveniles and cognitively vulnerable individuals (Leo, Costanzo, and Shaked-Schroer 2009; Kassin et al. 2014). In response, scholars

and practitioners have long advocated for a comparatively simple procedural safeguard, the electronic recording of custodial interrogations (Gangi 1984; Kassin et al. 2014). Yet, despite decades of advocacy and research supporting interrogation recording policies (Sullivan 2004; Kassin et al. 2014), the question of why so many state legislatures have resisted this reform remains underexplored.

Recording custodial interrogations creates a transparent and reviewable record of police-suspect interactions, allowing judges and jurors to evaluate better the voluntariness and reliability of confession evidence (Kassin et al. 2014; Sullivan 2004). By capturing not only what was said but also the conditions under which statements were obtained, recording requirements can deter coercive or psychologically manipulative interrogation practices (Kassin et al. 2014; Gangi 1984). Importantly, interrogation recordings are often understood as a reform that benefits all participants in the criminal justice system. For example, a complete audiovisual record enhances the reliability of evidence for prosecutors, protects law enforcement officers from unfounded allegations of misconduct, and safeguards suspects' due process rights by increasing transparency and accountability during interrogations (Leo and Richman 2007; Sullivan 2004). Taken together, this body of research demonstrates that recordings are a meaningful safeguard that strengthens the integrity of the interrogation process for all parties involved, making the resistance to their adoption all the more difficult to explain on evidentiary grounds alone.

Despite broad expert consensus and relatively low implementation costs, the adoption of interrogation recording laws has been uneven across US states (Bang et al. 2018; Kassin et al. 2014; Sullivan 2004). Since Illinois enacted the first statewide requirement in 2003, only twenty-three additional states have

1 William Lowthert is an MPP student at the McCourt School of Public Policy at Georgetown University

2 Christopher J. Normile, Ph.D. is an Assistant Professor of Psychology at Allegheny College

passed similar legislation, while five others have adopted recording mandates through state supreme court decisions. Twenty-one states still lack a statewide legal requirement (Deyoung 2014). Given the reform's demonstrated benefits, what explains why these twenty-one states have not acted? The answer lies not in any weakness in the evidence for recording but in the state-level partisan and racial dynamics that shape whether legislatures are willing to enact these reforms.

How Interrogation Recording Laws Are Enacted

Understanding how interrogation recording requirements have been adopted is essential to explaining why so many states have not adopted them. The pathway through which a recording mandate is enacted, whether through judicial action or legislative process, carries distinct political implications that shape the likelihood of adoption.

States that require the recording of custodial interrogations have adopted these policies through two primary pathways, judicial mandate and legislative action. These pathways differ not only in form but also in the degree to which they are exposed to partisan and electoral pressures, a distinction central to understanding why adoption has been so uneven across the nation. In some states, supreme courts have concluded that the failure to record custodial interrogations raises due-process or procedural concerns (Deyoung 2014; Sullivan 2004). The Alaska and Arkansas Supreme Courts, for example, held that an unrecorded custodial interrogation violates due process when recording is feasible, effectively conditioning the admissibility of custodial statements on whether an interrogation was recorded (Deyoung 2014; Sullivan 2004). Similarly, the supreme courts of Minnesota and New Jersey relied on their supervisory authority over trial courts to mandate the recording of interrogations through judicial precedent, emphasizing the role of recordings in promoting fairness and reliability in the criminal justice system (Deyoung 2014; Sullivan 2004). Other state courts have implemented interrogation recording requirements through rules of evidence or criminal procedure. In Utah and Indiana, for instance, courts adopted evidentiary rules that render unrecorded custodial statements

inadmissible in felony cases unless specific, enumerated exceptions apply (Deyoung 2014). What unites these judicial approaches is a critical institutional feature of courts being functionally insulated from partisan politics. This feature allows them to vindicate procedural protections and due process norms without facing the same electoral consequences as legislators, enabling the implementation of recording requirements even in states where legislatures have been unwilling or unable to act (Deyoung 2014; Garland 2005).

In addition to courts, many states have relied on the legislative process to enact laws requiring the recording of interrogations. Legislative adoption requires approval by both chambers of a state legislature and the governor, placing interrogation recordings squarely within partisan debates over crime, policing, and public safety. Unlike court-mandated reforms, legislative action is shaped by electoral accountability, partisan messaging, and interest group influence (Garland 2005; Yates and Fording 2005; Kent and Carmichael 2013). As a result, legislative consideration exposes interrogation recording laws to broader ideological conflict and increases the likelihood that reforms designed to strengthen due process will encounter political resistance (Garland 2005; Kent and Carmichael 2015; Carmichael and Kent 2017). This distinction between judicial insulation and legislative exposure helps explain the uneven adoption of interrogation recording laws and is a key point of the paper's broader argument. In states where courts have acted independently, recording mandates have advanced even under Republican-controlled legislatures. However, in states where adoption requires legislative action, partisan opposition and racialized framing of criminal justice policy have repeatedly stalled the reform.

Even without formal legislation, interrogation recording has become widespread in several states through voluntary compliance, court-driven incentives, or departmental policies. That recording can take hold through these alternative channels suggests that political resistance is directed at the act of codifying mandates rather than the practice of recording itself. In Rhode Island, for example, the legislature convened a task force to study the recording of custodial interrogations and to recommend appropriate policies and procedures. The task

force issued its final report in 2012, which prompted all police agencies in the state to adopt recording practices voluntarily (Bang et al. 2018; Innocence Project 2019; National Association of Criminal Defense Lawyers n.d.). Building on this effort, the Rhode Island Police Accreditation Commission implemented an interrogation recording policy in 2013, requiring the video recording of custodial interrogations of persons suspected of capital offenses. All forty-three police departments in the state agreed to the policy, resulting in what appears to be statewide compliance with recording serious felonies despite the absence of a statutory mandate. A similar outcome also occurred in Hawaii (Bang et al. 2018; Innocence Project 2019; National Association of Criminal Defense Lawyers n.d.).

Massachusetts reflects a dynamic shaped by judicial influence rather than legislative action. In *Commonwealth v. DiGiambattista* (2004), the Supreme Judicial Court held that when unrecorded custodial statements are introduced at trial, juries must be instructed that the court has expressed a preference for recording interrogations whenever practicable. To avoid these adverse jury instructions and the resulting credibility concerns, police departments throughout the state have adopted routine recording practices, effectively institutionalizing recording through evidentiary incentives rather than formal legislation (Sullivan 2004; Innocence Project 2019; National Association of Criminal Defense Lawyers n.d.). New Hampshire provides an additional example of judicially-driven implementation. In 2002, the New Hampshire Supreme Court held that a recorded custodial statement is admissible only if the entire post-Miranda interrogation was electronically recorded, thereby encouraging comprehensive recording practices through admissibility rules rather than statutory command (Sullivan 2004; Innocence Project 2019).

Race, Politics, and Partnerships

Although the benefits of recording custodial interrogations are well documented, the uneven adoption of these laws suggests that political factors are preventing their nationwide embrace. When these reforms enter the legislative arena, they become embedded in broader political debates over crime, policing, and public safety. These debates are not race-neutral. Nor are they

driven primarily by evidence. Instead, they are shaped by long-standing racial hierarchies and by how crime has been historically racialized in American politics (Chambliss 1994; Mendelberg 2001).

Racial prejudice, racial threat, and historical patterns of discrimination have influenced the development and enforcement of criminal law in the United States (Edelman 2006; Walker, Spohn, and DeLone 2018). During earlier periods of American history, criminal justice institutions often directly participated in or tacitly condoned racial violence. Although such overt practices have declined, their legacy continues to shape contemporary attitudes toward crime and punishment (Edelman 2006; Walker, Spohn, and DeLone 2018). Research consistently shows that racial prejudice and stereotypes are associated with greater public support for punitive criminal justice policies, particularly those that disproportionately affect minority communities (Cohn, Barkan, and Halteman 1991; Hurwitz and Peffley 1997). White Americans, who remain the majority population in most states, often hold implicit or explicit biases toward racial minorities, which influence perceptions of criminality and deservingness (Golden 2012; Pickett et al. 2012; Pickett et al. 2014). At the state level, these dynamics have been linked to higher incarceration rates, increased corrections spending, and more substantial support for capital punishment in jurisdictions with larger minority populations (Pritchard and Wiatr 2008; Stucky, Heimer, and Lang 2007; Jacobs, Malone, and Iles 2012). These findings point to a broader pattern in which the racial composition of a state shapes the political environment around criminal justice policy, creating conditions where reforms that strengthen due process protections face greater resistance.

Since interrogation recording laws are designed to strengthen procedural protections for criminal suspects, they can be perceived as politically sensitive in contexts where crime is racialized. In states where criminal suspects are implicitly associated with minority groups, reforms that enhance due process protections may be framed as favoring defendants over victims or as constraining law enforcement authority (Carmichael and Kent 2017; Chambliss 1994). As a result, interrogation recording laws do not fail on their merits but rather become

caught up in the racialized politics of crime, shaping both public opinion and the willingness of policymakers to advance due process reforms.

Racial context alone, however, does not fully explain resistance to interrogation recording laws. Its political effects are mediated through partisan competition and ideological commitments to punitive crime control (Garland 2005; Yates and Fording 2005; Carmichael and Kent 2014; Carmichael and Kent 2017). A large body of research finds that conservative political ideology is associated with stronger support for punitive approaches to crime, including harsh sentencing policies, capital punishment, and skepticism toward reforms perceived as benefiting criminal defendants (Moon et al. 2000; Vogel and Vogel 2003; Unnever and Cullen 2010).

These ideological preferences are reflected in state-level policymaking. Republican-dominated states have been found to adopt fewer policies aimed at reducing wrongful convictions or expanding procedural protections for defendants, while simultaneously maintaining higher incarceration rates and greater use of the death penalty (Jacobs et al. 2012; Jacobs and Carmichael 2001; Kent and Carmichael 2015). For interrogation recording laws specifically, this means that Republican-controlled legislatures are less likely to view recording mandates as common-sense safeguards and more likely to perceive them as constraints on police authority that shift procedural advantages toward suspects (Kent and Carmichael 2015; Carmichael and Kent 2017; Jacobs and Carmichael 2002). Even when proponents emphasize that recording

interrogations improves evidence quality and protects law enforcement officers from false allegations, these arguments may fail to overcome partisan narratives that prioritize toughness on crime and deference to police authority (Garland 2005; Carmichael and Kent 2017). This interaction between racial context and partisan control is what produces the patterns of adoption and resistance examined in the following section.

Patterns of Adoption Across States, 2003–2022

The political dynamics outlined above should be visible in how states have actually adopted or resisted interrogation recording laws. This section examines descriptive patterns in state adoption between 2003 and 2022, mapping where recording requirements exist, how they were adopted, and under what political conditions states have either enacted or resisted formal mandates.

As of 2022, twenty-four states have enacted interrogation recording requirements through legislative action, while five additional states require recording through state supreme court mandates. At the same time, twenty-one states, including the special cases discussed earlier, still lack a statewide statutory requirement to record custodial interrogations.

Among the twenty-four states that adopted interrogation recording laws through legislation, enactment occurred under a range of political conditions. Thirteen of these laws were passed when Democrats controlled both

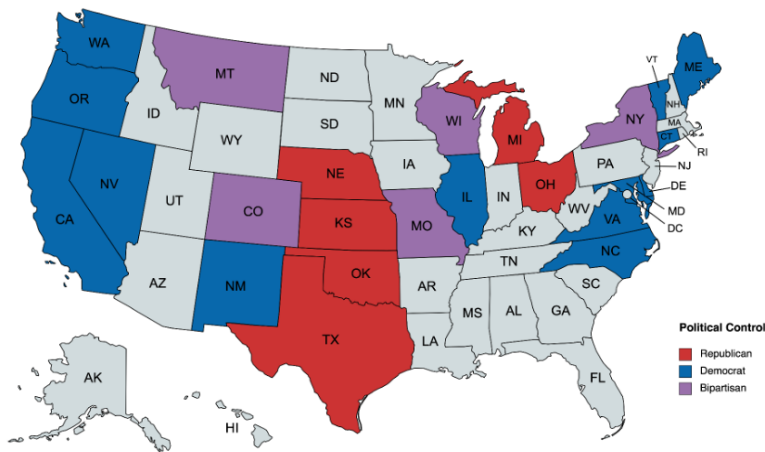


Figure 1: States That Have Passed Interrogation Recording Laws

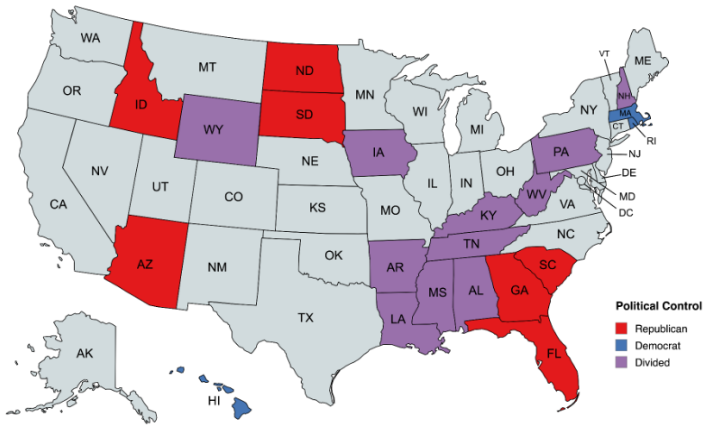


Figure 2: Post-2005 States That Have Not Passed Interrogation Recording Laws

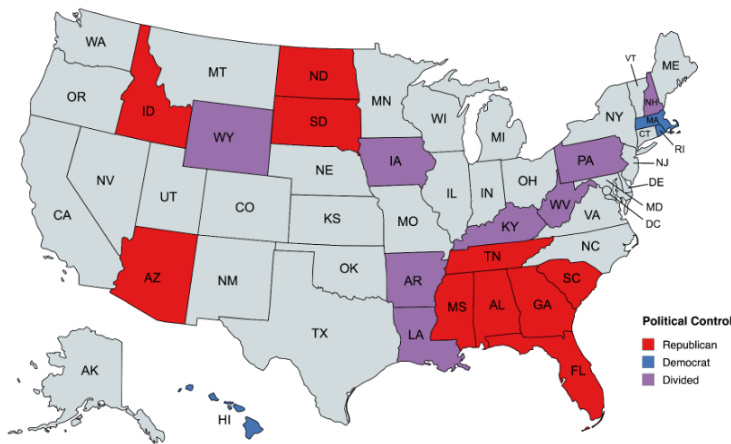


Figure 3: Post-2012 States That Have Not Passed Interrogation Recording Laws

legislative chambers and the governorship, six were enacted under Republican trifectas, and eight were adopted under divided government. These distributions indicate that interrogation recording laws are not exclusively the product of Democratic governance. However, Democratic-controlled states account for a clear majority of legislative adoptions, while Republican-controlled states represent a smaller share, suggesting that partisan control conditions the likelihood of legislative enactment. While both parties have passed these laws, Democratic-controlled states have done so at more than twice the rate of Republican-controlled ones.

A different pattern emerges among the twenty-one states that have not enacted interrogation recording laws through legislation. Beginning in 2005, seven of these non-adopting states experienced long-term Republican control, eleven were characterized

by divided or inconsistent partisan control, and only three experienced sustained Democratic control. States that have resisted adoption are far more likely to have experienced Republican or divided control than sustained Democratic governance.

Partisan differences become more pronounced when the analysis is restricted to the post-2012 period, during which most states adopted interrogation recording laws. Using this later time frame, ten of the twenty-one non-adopting states are classified as having long-term Republican legislative control, five are classified as divided, and only three remain under sustained Democratic control. This shift reflects the consolidation of Republican control in many state legislatures after 2010. It suggests that partisan dominance influences not only whether states adopt interrogation recording laws, but also whether adoption is delayed as

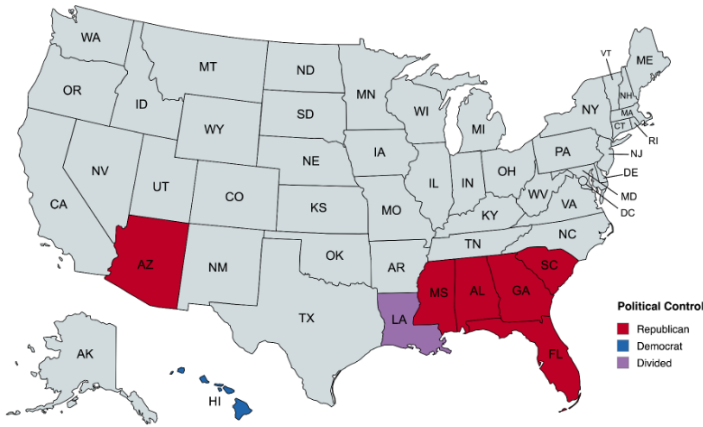


Figure 6: Post-2012 States That Have Not Passed Interrogation Recording Laws and Have A Substantial Minority Population

Among the eight racially diverse states that have not adopted interrogation recording laws, political control again plays a decisive role. In the post-2005 period, four experienced long-term Republican legislative control, three experienced bipartisan control, and only one experienced sustained Democratic control. Among racially diverse non-adopting states, Republican control is the most common political condition.

Why Evidence Is Not Enough: Expertise, Ideology, and Electoral Incentives

One explanation for the partisan divide observed in legislative adoption concerns differential trust in expert authority. A substantial body of research documents that Republican elites and voters are more skeptical of academic and scientific expertise than their Democratic counterparts, in part because universities and research institutions are perceived as ideologically liberal (Hamilton and Hargens 1993; Eagan et al. 2014; Langbert, Quain, and Klein 2016). These perceptions have contributed to broader Republican skepticism toward expert-driven policy recommendations, even in areas with strong empirical consensus. Although research indicates that experts’ liberal political leanings do not systematically bias their work (Reinero et al. 2020), these perceptions may nonetheless shape legislative receptivity to reforms championed by legal scholars and social scientists. This dynamic helps explain why decades of advocacy by legal psychologists and criminal justice experts have not translated into

uniform legislative adoption of interrogation recording laws.

Ideological commitments to punitive crime control further reinforce partisan resistance. Extensive research finds that Republican voters and elected officials are more supportive of policies emphasizing punishment, law enforcement discretion, and toughness on crime (Kent and Carmichael 2015). Within this framework, interrogation recording laws may be perceived as constraining police authority or shifting procedural advantages toward criminal suspects (Jacobs and Carmichael 2002).

Racial context further conditions these partisan dynamics. A large body of scholarship demonstrates that criminal justice policy in the United States has been deeply racialized, with political actors leveraging racialized fears to mobilize electoral support (Chambliss 1994; Mendelberg 2001; Weaver 2007). Republican strategies have historically emphasized street crime and disorder in ways that implicitly associate criminality with racial minority groups, appealing to white voters who perceive growing minority populations as threatening (Giles and Hertz 1994; Hurwitz and Peffley 2005; Jacobs and Tope 2008).

In this context, interrogation recording laws, which enhance procedural protections for criminal suspects, may be interpreted as reforms that empower populations already viewed through a racialized lens of suspicion. This dynamic helps explain why long-term Republican-controlled states with substantial minority populations appear especially resistant to interrogation recording laws.

Conclusion

The uneven adoption of interrogation recording laws illustrates that even widely-supported reforms do not spread on the strength of evidence alone. Recording custodial interrogations is a low-cost safeguard with demonstrated benefits for every party in the criminal justice system. Yet its adoption depends not on its merits but on the political environment in which it is considered.

This analysis shows that partisan control, particularly long-term Republican dominance, is closely associated with delayed or blocked legislative adoption, especially as the policy became more normalized after 2012. Racial context further shapes these outcomes by interacting with partisan incentives, reinforcing resistance in states where crime is politically racialized, and due process protections are framed as constraining law enforcement authority. The presence of widespread recording in several non-adopting states demonstrates that opposition often centers on the political costs of codifying mandates rather than on the practice itself. Courts, evidentiary rules, and voluntary departmental policies have played an important role in advancing recording where legislatures have been unwilling to act.

Expanding interrogation recording to the remaining twenty-one states will require strategies that account for these political realities. Advocates should focus on depoliticizing the reform by partnering with law enforcement organizations that already support recording, framing mandates around protecting officers and improving evidence quality, and pursuing judicial and administrative pathways that bypass legislative resistance. Closing this gap will not require additional research to prove that recording works. It will require making the reform politically viable in the states that have resisted it the longest.

References

- Bang, Brandon L, Duane Stanton, Craig Hemmens, and Mary K Stohr. 2018. "Police Recording of Custodial Interrogations." *International Journal of Police Science & Management* 20 (1): 3–18. <https://doi.org/10.1177/1461355717750172>.
- Berlinski, Nicolas, Margaret Doyle, Andrew M. Guess, Gabrielle Levy, Benjamin Lyons, Jacob M. Montgomery, Brendan Nyhan, and Jason Reifler. 2021. "The Effects of Unsubstantiated Claims of Voter Fraud on Confidence in Elections." *Journal of Experimental Political Science* 10 (1): 34–49. <https://doi.org/10.1017/xps.2021.18>.
- Bureau of Justice Statistics. 2023. "Prisoners In 2022 – Statistical Tables | Bureau Of Justice Statistics." *Bjs.Ojp.Gov*. <https://bjs.ojp.gov/press-release/prisoners-2022-statistical-tables> (January 31, 2024).
- Carmichael, Jason, and Stephanie Kent. 2014. "The Persistent Significance of Racial and Economic Inequality on the Size of Municipal Police Forces in the United States, 1980–2010." *Social Problems* 61 (2): 259–82. <https://doi.org/10.1525/sp.2014.12213>.
- Carmichael, Jason T., and Stephanie L. Kent. 2017. "The Racial Politics of Due Process Protection." *Criminal Justice Review* 42 (1): 58–76. <https://doi.org/10.1177/0734016816684925>.
- Chambliss, William J. 1994. "Policing the Ghetto Underclass: The Politics of Law and Law Enforcement." *Social Problems* 41 (2): 177–94. <https://doi.org/10.2307/3096929>.
- Cohn, Steven F., Steven E. Barkan, and William A. Halteman. 1991. "Punitive Attitudes toward Criminals: Racial Consensus or Racial Conflict?" *Social Problems* 38 (2): 287–96. <https://doi.org/10.2307/800534>.
- Deyoung, Jacob. 2014. "Rule 4.7 of the Arkansas Rules of Criminal Procedure: A Squandered Opportunity or a Step in the Right Direction?" <https://cpb-us-e1.wpmucdn.com/wordpressua.uark.edu/dist/0/285/files/2014/06/ArkLRev-67-1-DeYoung.pdf>.
- Eagan, Kevin, Ellen Bara, Stolzenberg Jennifer, Berdan Lozano, Melissa Aragon, Maria Suchard, and Sylvia Hurtado. 2014. "Undergraduate Teaching Faculty: The 2013–2014 HERI Faculty Survey." <https://heri.ucla.edu/monographs/HERI-FAC2014-monograph.pdf>.
- Edelman, Bryan C. 2006. *Racial Prejudice, Juror Empathy, and Sentencing in Death Penalty Cases*. LFB Scholarly Publishing.
- Gangi, William. 1984. "Inbau-Kamisar Debate: Time for Round 2." *W. St. UL Rev. HeinOnline*. https://www.academia.edu/379910/Inbau_Kamisar_Debate_Time_for_Round_2_The.
- Garland, David. 2005. "Capital Punishment and American Culture." *Punishment & Society* 7 (4): 347–76. <https://doi.org/10.1177/1462474505057097>.
- Giles, Micheal W., and Kaenan Hertz. 1994. "Racial Threat and Partisan Identification." *American Political Science Review* 88 (2): 317–26. <https://doi.org/10.2307/2944706>.
- Golden, Kristen Brown. 2021. "White Privilege: Unconscious Racism, Freud, and Neuroscience of Implicit Bias." *Critical Philosophy of Race* 9 (2): 295. <https://doi.org/10.5325/critphilrace.9.2.0295>.
- Golden, Kristin M. 2012. "The Effect of Static, Dynamic, and Perceptual Measures of Minority Threat on Fear of Crime." *Journal of Ethnicity in Criminal Justice* 10 (2): 108–28. <https://doi.org/10.1080/15377938.2012.669655>.
- Hamilton, Richard F., and Lowell L. Hargens. 1993. "The Politics of the Professors: Self-Identifications, 1969–1984." *Social Forces* 71 (3): 603. <https://doi.org/10.2307/2579887>.
- Hurwitz, J., and M. Peffley. 2005. "Playing the Race Card in the Post-Willie Horton Era: The Impact of Racialized Code Words on Support for Punitive Crime Policy." *Public Opinion Quarterly* 69 (1): 99–112. <https://doi.org/10.1093/poq/nfi004>.

- Hurwitz, Jon, and Mark Peffley. 1997. "Public Perceptions of Race and Crime: The Role of Racial Stereotypes." *American Journal of Political Science* 41 (2): 375–401. <https://doi.org/10.2307/2111769>.
- Innocence Project. 2023. "Recording of Interrogations in New Hampshire - Innocence Project." Innocence Project. April 28, 2023. <https://innocenceproject.org/policies/recording-of-interrogations-in-new-hampshire/>.
- Jacobs, D., and J. T. Carmichael. 2001. "The Politics of Punishment across Time and Space: A Pooled Time-Series Analysis of Imprisonment Rates." *Social Forces* 80 (1): 61–89. <https://doi.org/10.1353/sof.2001.0070>.
- Jacobs, David, and Daniel Tope. 2008. "Race, Crime, and Republican Strength: Minority Politics in the Post-Civil Rights Era." *Social Science Research* 37 (4): 1116–29. <https://doi.org/10.1016/j.ssresearch.2008.05.004>.
- Jacobs, David, and Jason T. Carmichael. 2002. "The Political Sociology of the Death Penalty: A Pooled Time-Series Analysis." *American Sociological Review* 67 (1): 109. <https://doi.org/10.2307/3088936>.
- Jacobs, David, and Ronald E. Helms. 1996. "Toward a Political Model of Incarceration: A Time-Series Examination of Multiple Explanations for Prison Admission Rates." *American Journal of Sociology* 102 (2): 323–57. <https://doi.org/10.1086/230949>.
- Jacobs, David, Chad Malone, and Gale Iles. 2012. "Race and Imprisonments: Vigilante Violence, Minority Threat, and Racial Politics." *The Sociological Quarterly* 53 (2): 166–87. <https://doi.org/10.1111/j.1533-8525.2012.01230.x>.
- Jacobs, David, Zhenchao Qian, Jason T. Carmichael, and Stephanie L. Kent. 2007. "Who Survives on Death Row? An Individual and Contextual Analysis." *American Sociological Review* 72 (4): 610–32. <https://doi.org/10.1177/000312240707200406>.
- Kassin, Saul M., Jeff Kukucka, Victoria Z. Lawson, and John DeCarlo. 2014. "Does Video Recording Alter the Behavior of Police during Interrogation? A Mock Crime-And-Investigation Study." *Law and Human Behavior* 38 (1): 73–83. <https://doi.org/10.1037/lhb0000047>.
- Kent, Stephanie L., and Jason T. Carmichael. 2013. "Racial Residential Segregation and Social Control: A Panel Study of the Variation in Police Strength across U.S. Cities, 1980–2010." *American Journal of Criminal Justice* 39 (2): 228–49. <https://doi.org/10.1007/s12103-013-9212-8>.
- Kent, Stephanie L., and Jason T. Carmichael. 2015. "Legislative Responses to Wrongful Conviction: Do Partisan Principals and Advocacy Efforts Influence State-Level Criminal Justice Policy?" *Social Science Research* 52 (July): 147–60. <https://doi.org/10.1016/j.ssresearch.2015.01.004>.
- Langbert, Mitchell, Anthony Quain, and Daniel Klein. 2016. "Faculty Voter Registration in Economics, History, Journalism, Law, and Psychology." https://econjwatch.org/file_download/944/LangbertQuainKlein-Sept2016.pdf?mimetype=pdf.
- Leo, Richard A., and Kimberly D. Richman. 2007. "Mandate The Electronic Recording Of Police Interrogations." *Criminology & Public Policy* 6 (4): 791–98. <https://doi.org/10.1111/j.1745-9133.2007.00467.x>.
- Leo, Richard A., Mark Costanzo, and Netta Shaked-Schroer. 2009. "Psychological and Cultural Aspects of Interrogations and False Confessions: Using Research to Inform Legal Decision-Making." *Papers.ssrn.com*. Rochester, NY. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1537331.
- Mendelberg, Tali. 2001. *The Race Card: Campaign Strategy, Implicit Messages, and the Norm of Equality*. Princeton, N.J.; Oxford: Princeton University Press.

- Moon, Melissa M., Jody L. Sundt, Francis T. Cullen, and John Paul Wright. 2000. "Is Child Saving Dead? Public Support for Juvenile Rehabilitation." *Crime & Delinquency* 46 (1): 38–60. <https://doi.org/10.1177/0011128700046001003>.
- Moon, Melissa M., John Paul Wright, Francis T. Cullen, and Jennifer A. Pealer. 2000. "Putting Kids to Death: Specifying Public Support for Juvenile Capital Punishment." *Justice Quarterly* 17 (4): 663–84. <https://doi.org/10.1080/07418820000094711>.
- Motta, Matthew. 2017. "The Dynamics and Political Implications of Anti-Intellectualism in the United States." *American Politics Research* 46 (3): 465–98. <https://doi.org/10.1177/1532673x17719507>.
- National Association of Criminal Defense Lawyers. n.d. "NACDL - Recording Interrogations." NACDL - National Association of Criminal Defense Lawyers. <https://www.nacdl.org/Landing/RecordingInterrogations>.
- Pickett, Justin T., Kelly Welch, Ted Chiricos, and Marc Gertz. 2014. "Racial Crime Stereotypes and Offender Juvenility." *Race and Justice* 4 (4): 381–405. <https://doi.org/10.1177/2153368714542007>.
- Pickett, Justin T., Ted Chiricos, Kristin M. Golden, and Marc Gertz. 2012. "Reconsidering The Relationship Between Perceived Neighborhood Racial Composition And Whites' Perceptions Of Victimization Risk: Do Racial Stereotypes Matter?" *Criminology* 50 (1): 145–86. <https://doi.org/10.1111/j.1745-9125.2011.00255.x>.
- Pritchard, Anita, and Michael Wiatrowski. 2008. "Race and Capital Punishment: State Level Analysis of the Effects of Race on States' Capital Punishment Policies." *Journal of Ethnicity in Criminal Justice* 6 (2): 103–21. <https://doi.org/10.1080/15377930802096488>.
- Reinero, Diego A., Julian A. Wills, William J. Brady, Peter Mende-Siedlecki, Jarret T. Crawford, and Jay J. Van Bavel. 2020. "Is the Political Slant of Psychology Research Related to Scientific Replicability?" *Perspectives on Psychological Science* 15 (6): 1310–28. <https://doi.org/10.1177/1745691620924463>.
- Stucky, Thomas D. 2011. "The Conditional Effects of Race and Politics on Social Control." *Journal of Research in Crime and Delinquency* 49 (1): 3–30. <https://doi.org/10.1177/0022427810393020>.
- Stucky, Thomas D., Karen Heimer, and Joseph B. Lang. 2007. "A Bigger Piece of the Pie? State Corrections Spending and the Politics of Social Order." *Journal of Research in Crime and Delinquency* 44 (1): 91–123. <https://doi.org/10.1177/0022427806295617>.
- Sullivan, Thomas. 2004. "Electronic Recording of Custodial Interrogations: Everybody Wins, 95." *J. Crim. L. & Criminology* 1127. <https://scholarlycommons.law.northwestern.edu/cgi/viewcontent.cgi?article=7203&context=jclc>.
- The Innocence Project. 2019. "Innocence Project: Transforming Systems." Innocence Project. October 29, 2019. <https://innocenceproject.org/transforming-systems/>.
- Udani, Adriano, David C. Kimball, and Brian Fogarty. 2018. "How Local Media Coverage of Voter Fraud Influences Partisan Perceptions in the United States." *State Politics & Policy Quarterly* 18 (2): 193–210. <https://doi.org/10.1177/1532440018766907>.
- Unnever, James D., and Francis T. Cullen. 2010. "The Social Sources Of Americans' Punitiveness: A Test Of Three Competing Models." *Criminology* 48 (1): 99–129. <https://doi.org/10.1111/j.1745-9125.2010.00181.x>.
- Vogel, Brenda L, and Ronald E Vogel. 2003. "The Age of Death: Appraising Public Opinion of Juvenile Capital Punishment." *Journal of Criminal Justice* 31 (2): 169–83. [https://doi.org/10.1016/s0047-2352\(02\)00223-4](https://doi.org/10.1016/s0047-2352(02)00223-4).

- Walker, Samuel, Cassia Spohn, and Miriam Delone. 2018. *The Color of Justice: Race, Ethnicity, and Crime in America*. 6th ed. Boston, MA: Cengage Learning.
- Weaver, Vesla M. 2007. "Frontlash: Race and the Development of Punitive Crime Policy." *Studies in American Political Development* 21 (2): 230–65. <https://doi.org/10.1017/s0898588x07000211>.
- Yates, Jeff, and Richard Fording. 2005. "Politics and State Punitiveness in Black and White." *The Journal of Politics* 67 (4): 1099–1121. <https://doi.org/10.1111/j.1468-2508.2005.00352.x>.
- Zalman, Marvin, Brad Smith, and Angie Kiger. 2008. "Officials' Estimates of the Incidence of 'Actual Innocence' Convictions." *Justice Quarterly* 25 (1): 72–100. <https://doi.org/10.1080/07418820801954563>.
- Zalman, Marvin, Laura L. Rubino, and Brad Smith. 2017. "Beyond Police Compliance with Electronic Recording of Interrogation Legislation: Toward Error Reduction." *Criminal Justice Policy Review* 30 (4): 627–55. <https://doi.org/10.1177/0887403417718241>.

Climate Change Adaptation Strategies and Care Ethics in Chicago, Illinois

Megan Morrell¹

Abstract

In response to the deleterious effects of climate change, many municipalities around the world have begun fortifying green spaces as climate change adaptation strategies. One curious case comes from the city of Chicago, Illinois, which has developed and funded a program called “Our Roots Chicago” to boost access to green spaces throughout the city by planting more trees on the South and West Sides. The North Side of Chicago, which contains predominantly wealthy white communities, has a much denser tree canopy than the South and West Sides, which are predominantly Black and Latine, and historically disinvested. The South and West Sides also experience some of the worst effects of climate change, including higher surface temperatures and more frequent flooding. While it may initially appear that planting more trees could only benefit the South and West sides, there are some important ethical complexities and limitations for this strategy. In this article, I utilize a feminist-Indigenous Care Ethics framework to analyze this policy. The benefits of Our Roots Chicago include reductions in surface temperature, sequestration of carbon and pollutants, improvements to mental and physical health, and opportunities to heal social and relational dysfunctions. Nevertheless, the limitations of this strategy include considerations of the ecosystem’s natural biome, green gentrification, and its colonial entanglements. I ultimately find that a Care Ethics framework is a powerful tool to guide inquiry and decision making for future climate change mitigation policies like Our Roots Chicago, which pose complex effects to communities.

Introduction

Our Roots Chicago

The City of Chicago designed the “Our Roots Chicago” program in 2022 after an influx of federal funding from the American Rescue Plan Act in the wake of the COVID-19 Pandemic and the resultant economic downturn. These federal funds were channeled into the Chicago Rescue Plan (City of Chicago n.d.c), aiming to use this “once-in-a-generation federal funding to create an equity-based investment strategy to catalyze a sustainable economic recovery from the COVID-19 pandemic” (City of Chicago n.d.a). Our Roots Chicago was thus constructed under the City’s Bureau of Forestry with the goal to plant 75,000 trees in the city in five years, with primary consideration for existing inequities and environmental injustices throughout Chicago, which have led the Far South and West Sides to have lower tree canopy density (City of Chicago n.d.b).

Tree canopy disparities around Chicago reflect the city’s enduring history of racial segregation. Redlining, zoning discrimination, and racial covenants amongst other legal, political, and social techniques of separation and

domination continue to shape the city’s physical and social environment (Kim et al. 2024; Faber et al. 2024). Alongside policies aimed at segregating Chicagoans, investment funneled towards predominantly white neighborhoods in the city, leading to disparities in many types of public infrastructure, including trees and greenery. Furthermore, areas in the city that are zoned for industrial uses are also predominantly in Black and Brown communities and have tended to contain fewer intentional green spaces because of the zoning. One method of measuring green space in a city is through tree canopies, which mirror the various social, health, and wealth inequities represented in segregated Chicago (see Appendix A). Because of these inequities, Our Roots Chicago has focused its planting on underserved neighborhoods on the South and West Sides, with the allocation of trees being conducted based on existing tree canopy measures, aiming to reach greater equity across Chicago’s neighborhoods (see Appendix B).

Care Ethics Framework

Care Ethics denote a constellation of ethical values and perspectives that emphasize the

¹ Megan Morrell is an MA in Social Sciences Research student at the University of Chicago

importance of interconnectivity and care for others, which can be a useful framework to assess the ethical implications of Our Roots Chicago. Scholars Kyle Powys Whyte (Potawatomi) and Chris Cuomo (unmarked)² are particularly salient voices in the Care Ethical tradition, focusing their work on both feminist and Indigenous perspectives, which “offer a range of related ideas and tools for environmental ethics [because they] delve into deep connections and moral commitments between nonhumans and humans to guide ethical forms of environmental decision making and environmental science” (Whyte et al. 2016, 1). In the case of Our Roots Chicago, Care Ethics can guide the assessment of this strategy for its ethical standing. Furthermore, Care Ethics in the Indigenous and feminist traditions both describe “approaches to moral life and community that are grounded in virtues, practices, and knowledges associated with appropriate caring and caretaking of self and others” (Whyte et al. 2016, 2). Notably, this approach contrasts many Western paradigmatic ethical theories that praise moral, impersonal, and abstract ethical calculations rather than the affective dimensions of morality (Whyte et al. 2016, 2). Care Ethics embraces the reality of deep dependence and interdependence between human and non-human beings and the importance of caretaking and forming close relational bonds (Whyte et al. 2016, 2; Gilligan 1993).

Methodology

This policy review will operationalize a Care Ethics guiding framework in an analysis of Our Roots Chicago that is both pragmatically and ethically rooted. Our Roots Chicago explicitly cites a philosophical grounding in environmental justice and should thus be assessed on its ethical grounds. Kincentricity is a key criterion for Care Ethics, viewing humans and the world as deeply interconnected and viewing non-human beings and the Earth as having ethical weight (Martinez et al. 2023). Humans remain important in a kincentric framework and are considered within a larger ecosystem

and relation of ethical subjects. Kincentrism overlaps greatly with the term Ecocentrism but is grounded in traditions of Indigenous thought throughout the world. Anthropocentrism, alternatively, positions humans at the center of ethical concern and argues that humans are the only beings deserving of ethical consideration. In this policy analysis, the extent to which the policy harms, benefits, considers, or ignores the needs of the local ecosystem and Chicagoans will be evaluated. Ultimately, I find that while the city’s commitments are indeed an exciting step towards healing the environmental and social dysfunctions in the city that disproportionately harm marginalized communities, a vision is missing regarding what a deep caretaking for all beings, human and non-human, could look like.

Analysis

Benefits of Tree Canopy Initiatives:

The benefits of augmenting green spaces, especially through the planting of trees, in the Chicago case are robust, and surround efficacy in reducing urban surface temperature, sequestering carbon and helping to reduce the pollutants responsible for climate change and beyond, the benefits to human mental and physical health, and the potential to restore destroyed natural biomes. Many cities experience what is called the ‘heat island effect,’ where cities feel hotter temperatures than their rural surroundings because surfaces such as concrete or steel absorb more heat than trees or vegetation (EPA 2024). First, tree canopies reduce surface temperatures through the interception of solar radiation through their leaves, which provide shade for the areas below them, which prevents the underneath surface from absorbing shortwave radiation, leading to less heat (NASA n.d.). For example, one may notice that in the summer it is much cooler in the shade, and this is because the radiation is being blocked from hitting one’s skin. Moreover, trees and other plants also evapotranspire, which allows them to absorb solar radiation and release water, which reduces the temperature of the surrounding area. In sum, these processes

2 While it is common practice in contemporary Indigenous scholarship to include acknowledgement of tribal affiliation, Max Liboiron (Metis/Mechif) illuminates the ways that leaving non-Indigenous scholars without a relational acknowledgement centers settlers and whiteness as the norm. Joining this practice of introduction, I include parentheses with any self-identified tribal affiliation or the term ‘unmarked’ to acknowledge the various relationships to land and state that are held by the cited individuals. This is cited in the first usage of a scholar’s name. For more details and discussion, see Liboiron, *Pollution is Colonialism*.

“combined with oasis and clothesline effects, even a single tree can moderate the microclimate, whereas large parks can extend the effects to the surrounding built environment” (Rahman et al. 2017). This phenomenon is one of the primary reasons cited for planting trees and other vegetation in green spaces and allows for communities to maintain cooler temperatures and sustain in a warming climate.

The planting of trees in Our Roots Chicago can also reduce the risk of flooding after heavy rainfall or severe weather events, which is becoming increasingly problematic for many communities in Chicago as our climate changes. Chatham and Calumet City, located on the far South Side of the Chicagoland area, are spaces that are particularly flood-prone. Trees and other vegetation work to “reduce rainwater runoff by intercepting rainfall in the canopy, evaporating water from the leaves, enhancing the infiltration around the trees’ roots, and storing water in trees’ trunks” (Qin 2020). Likewise, the anti-flooding properties innate in trees and other green infrastructure allow for greater adaptation for a community’s resilience in a climate that is likely to experience increased extreme weather events and shifts in both temperatures and precipitation (Intergovernmental Panel on Climate Change 2021).

Trees and other vegetation also sequester carbon dioxide, which is a major contributor to the greenhouse effect driving global climate change. The greenhouse effect describes the process by which thermal radiation from the sun is absorbed by Earth’s surface and retained by a layer of gasses such as carbon dioxide and methane. As the presence of greenhouse gases in the atmosphere increases, the Earth’s surface warms (IPCC 2024). The capacity of trees to sequester carbon increases as the tree grows, as “it stores more carbon by holding it in its accumulated tissue [leading to] the amount of carbon annually sequestered [increasing] with the size and health of the trees” (National Park Service 2022). Other non-climate related pollutants are also mitigated by vegetation, including trees such as ozone, nitrogen dioxide, sulfur dioxide, carbon monoxide, and particulate matter from the atmosphere (City of Chicago 2023).

Thus, trees reduce some of the scientific and biological concerns of urban landscapes in the face of climate change, and thus far appear to

complement the principles of interconnectivity in Care Ethics. The reduction in temperature, pollution, and flooding benefits the whole Chicago ecosystem’s adaptability in the face of a shifting climate. From a Care Ethics perspective, adaptation to climate changes is certainly a positive outcome for this program, supporting both humans and their non-human relations. Whyte (2014) often emphasizes the importance of “collective continuance,” as a society’s adaptive capacity or ability to continue through challenging scenarios such as climate change. The ability to persist through the challenges of climate change is an ultimate goal for Care Ethics, positioning Our Roots Chicago favorably for this ethical paradigm.

Living in proximity to green spaces in cities can also have positive effects on human mental and physical health through the promotion of physical activity, which in turn can have benefits for mental health (Turner-Skoff et al. 2019). Parks, for instance, often contain the infrastructure necessary for many sports and activities, (i.e. a basketball court, bike trail, or playground), which make physical exercise more accessible. Further, the concept of ecosystem service ecotherapy, or “the ability of interaction with nature to enhance healing and growth” is augmented in the presence of greater access to green spaces (Summers and Vivian 2018). Ecotherapy can take the form of activities such as “social and therapeutic horticulture, animal-assisted interventions, care farming, green exercise, environmental conservation and wilderness therapy” and can aid in the healing of mental health concerns (Chaudhury and Banerjee 2020).

Ultimately, in regard to human health, “trees not only make people happier and healthier, but they make communities more livable” (Chaudhury and Banerjee 2020). Care for the body and mind is a central feature of Care Ethics, making this dimension of tree planting a clear benefit. Furthermore, plants also can assist in the health and growth of other nearby plants, also called companion planting or plant associations. While many plants compete with each other for resources such as sunlight, water, and nutrients, many species complement each other and grow synergistically. Care Ethics’ intentional ecocentric and kincentric perspective makes the consideration of non-human benefits of increased green space

possible whereas environmental justice tends to ignore these dimensions.

Green spaces can also provide areas for greater social connection and cohesion (Clarke et al. 2023). Because green spaces serve as locations that are often free and able to host gatherings of people, they are embedded in the social life of the city. Social cohesion is a particularly useful concept to understand how green spaces work upon social life, as the “extent of connectedness and solidarity among groups in society” (Manca 2014). In terms of climate change, Kyle Powys Whyte argues that there are two major tipping point categories with which climate ethics ought to contend: an ecological tipping point, which is concerned with “the inaction of societies to mitigate their contributions to atmospheric concentrations of greenhouse gasses threaten to have irreversible and dangerous effects,” and a relational tipping point, concerned with “the inaction of societies to establish or maintain relational qualities connecting societal institutions together for the sake of coordinated action” (Whyte 2020). For Whyte, the relational tipping point has already been passed, especially from the perspective of many Indigenous peoples, who “approach climate change already having been through transformations of their societies induced by colonial violence” (Whyte 2018). Indigenous communities in many ways “have seen the end of their respective worlds’ [and] survived the apocalypse” (Gross 2016). The toxic relational dynamics that have fueled climate change need to be addressed as the primary concern for Whyte, and creating physical spaces for community-building can facilitate this important healing that is centered around care for one another.

Limitations of Our Roots

While there are a variety of benefits for green spaces and increased vegetation in urban contexts, there are also many limitations: the cost of planting greenery and upkeep, considerations about the urban setting’s natural biome, the role that vegetation plays into systems of gentrification and urban renewal, and the emphasis on this climate mitigation strategy. Regarding the cost of planting and upkeep green space, Our Roots Chicago has a total budget of \$46 million to be spent over 5 years (Bloomberg 2024). The City of Chicago also dedicated \$21,660,736 per year on tree

maintenance in 2023 (City of Chicago 2023). While this spending is indeed promising for providing benefits for Chicagoans, there are always opportunity costs associated with this spending, where money can be spent on other essential projects that can care for Chicagoans and the Earth in other, different ways.

It is also vital to consider whether the vegetation being planted is biologically appropriate for the ecosystemic context in which it is being planted. For instance, “Chicago straddles two USDA plant hardiness zones (5b in the north and 6a in the south) and is predicted to shift toward complete 6a designation by 2040” (United States Department of Agriculture 2021). According to the USDA, the Plant Hardiness Zone Map is based on the average annual extreme minimum winter temperature, displayed in 13 zones with half zones designated as ‘a’ and ‘b’ (United States Department of Agriculture n.d.a). In a changing climate, “the city’s plant hardiness zones are predicted to continue to warm, reaching 6b under low-emissions scenarios and 7a under high emissions scenarios by 2100” (Matthews 2018). Climate change has thus redefined the types of plants that can thrive in Chicago’s climate in favor of species that can adapt to drier and warmer conditions. Chicago’s natural ecosystems are quite diverse, especially examining the larger Cook County scale, which includes wetlands, prairies, savannahs, and woodlands (Forest Preserves of Cook County 2022). The heterogeneity of Chicago’s ecosystems calls into question Our Roots Chicago’s decision to focus on restoring one aspect of the natural ecosystems- trees- and not other features such as native grasses or shrubs that can provide many of the same biological benefits, though less shade. A Care Ethics lens challenges whether tree planting rather than the planting of native grasses or other vegetation cares for the health of the natural biome – a kincentric rather than anthropocentric question. Alternatively, environmental justice can tend to focus on human benefits of equitable access to a healthy environment; incredibly important, yet insufficient in the larger picture of restoring aspects of the native ecosystems (including wetlands and grasslands) in Chicago that have been destroyed in the process of urbanization.

Trees also take a long time to mature, challenging the ethics of shaping Chicago’s vegetation landscape for future generations. The

burr oak, *Quercus macrocarpa*, for instance, is a commonly found tree that is native to the Chicagoland area and can take anywhere from 10-35 years before it matures and begins to produce acorns for reproduction. (United States Department of Agriculture n.d.b.). Caring for future generations and their abilities to live in a habitable ecosystem is a foundational concept in Our Roots Chicago and thus makes this dimension less of a benefit or limitation, but more of a complication of any climate change mitigation action at all. More consideration for this complication is necessary but goes beyond the boundaries of this article.

Trees and other green spaces are also entangled in socio-political systems of domination such as gentrification (also called green gentrification), and urban renewal projects. Gentrification describes “the process by which under-resourced neighborhoods are developed and experience a migration of affluent newcomers,” and neighborhood development can include the expansion of green spaces (Jelks et al. 2021). The 606, a greenway trail that opened in 2015 and stretches through Chicago’s West Side neighborhoods of Logan Square, Wicker Park, Bucktown and Humboldt Park, is one curious location of intersecting motions of gentrification, segregation, and green spaces in Chicago. Harris et al. (2020) describe how this 606 greenway highlights the urban park paradox, “where green space provides benefits to communities while simultaneously generating unintentional consequences that potentially reinforce segregation and social inequities.” Many Latine communities in Humboldt Park expressed not feeling ownership over the shared space, and were not consulted before its construction (Rigolon et al. 2020). This area of the city continues to become more expensive for its residents (Francis n.d.), fueling anti-gentrification movements in the area in response to the trail (Parrella-Aureli 2024). Thus, the creation of green spaces calls into question a myriad of costs and benefits that complicate the ethics of simply planting trees in underserved neighborhoods without explicit and robust gentrification mitigating actions.

Furthermore, there is also a larger concern about the impact of green space development and increases in the cost of living in a neighborhood (especially rent prices), which can contribute to the displacement of low-

income and working-class populations (Lee et al. 2023). Without intentional safeguards, such as rent control or a universal dedication to housing all people, the very populations that Our Roots Chicago and other green space initiatives seek to serve may be harmed by the magnification of socioeconomic disparities. While there are programs conducted by the City of Chicago that focus on mitigating gentrification (City of Chicago 2021), the fragmentary process by which Our Roots Chicago recognizes the interconnections of social systems of housing and movement alongside the construction of Chicago’s urban environment leaves already marginalized populations vulnerable to further harm. This process would reverse the potential benefits for marginalized populations outlined in the benefits section and therefore would not result in care.

It is also important to recognize that this municipal tree planting initiative is always operating under a baseline state of colonial occupation of the land now known as Chicago. The Potawatomi, Odawa, Ojibwe, Miami, Ho Chunk, Menominee, Sac, and Fox nations continue to have legitimate political claims to the lands that Chicago resides on. These communities have called this area home since time immemorial, developing inseparable ties to the land and deep knowledge about its ecosystems. For Whyte (2020), healing the relational rupture that is foundational to ongoing colonialism is vital before any other steps are taken in order to reduce the risk that climate change mitigation techniques or solutions do not further harm Indigenous peoples. Whyte goes as far to say that urgent climate mitigation and solutions should be deprioritized, and our attention should instead be focused on relational healing. Paying land and monetary reparations back to Indigenous peoples is neglected in this course of action in Our Roots Chicago and all of the City of Chicago’s environmental and social programs. After all, the goal of planting 75,000 trees in the next 5 years may or may not align with current and traditional Indigenous land management techniques and knowledges, which highlights the necessity of involving more explicitly Indigenous communities in climate mitigation plans such as Our Roots Chicago.

Discussion

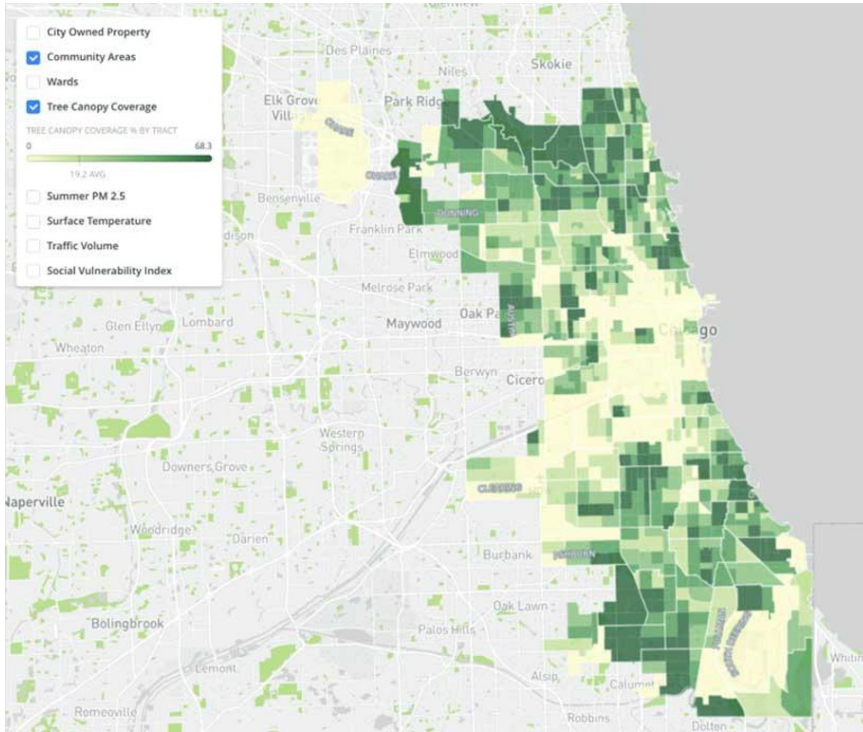
Ultimately, Our Roots Chicago presents a multidimensional ethical case study of many

potential benefits for humans and non-human beings, with some consequential limiting factors. Regarding the appropriateness of trees as the choice of Our Roots Chicago, there ought to be further consideration of other vegetation types that may fit the ecological needs of this landscape best. I do not believe this limitation is existentially problematic for the program, but a necessary contention under Care Ethics. Similarly, the costs of this program and subsequent opportunity cost are a dimension that necessitate further investigation. Our Roots Chicago has proven itself to have many compelling benefits (with a few important limitations), and thus worthy of funding. The amount of funding and its sources is an important topic for future investigations.

Many of the issues with planting more trees through this program have more to do with socio-economic systems at play than with the act itself (given they are biologically appropriate to the Chicago ecosystem). For instance, planting trees in Chicago's neighborhoods that have been historically deprived of green investments is only harmful in a context of property values and land developers with profit motivations that lead to the displacement of communities. The

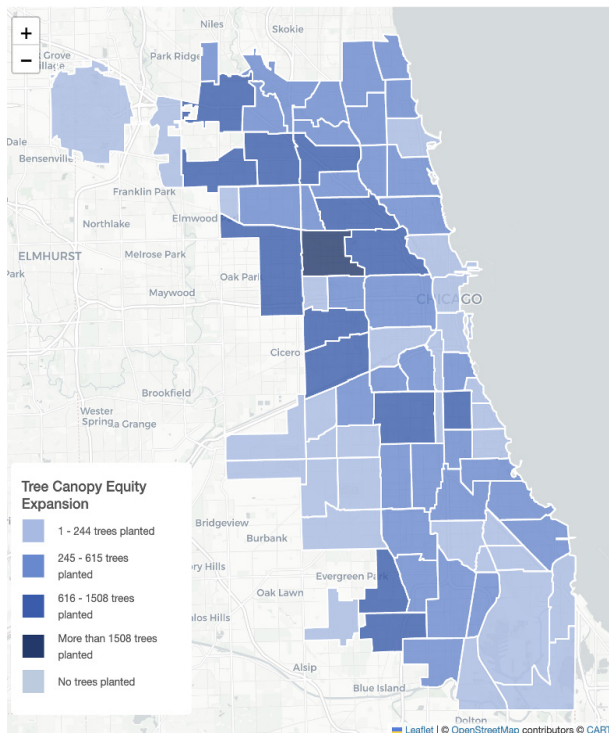
system of racial capitalism is the subterranean process at play that makes gentrification and displacement possible, not the trees themselves (Rucks-Ahidiana 2022). Healing colonial relations and creating infrastructure that ensures housing instead of green gentrification is also a part of a larger project of Care Ethics that is only somewhat recognized by environmental justice frameworks that are often narrowly anthropocentric and lack recognition of interconnected systems of environmental and social domination. Ultimately, a Care Ethical framework can help us to imagine beyond our current political struggles towards a future of greater equity and systems that can allow us to live in reparative relations with Chicago's landscape and peoples beyond the confines of social, economic, and environmental domination.

Appendix



Appendix Figure A:

This map shows tree canopy coverage as % of trees by census tract. (Witynski, Max. 2021)



Appendix Figure B:

Tree Canopy Equity Expansion measured by number of trees planted per neighborhood area (City of Chicago. n.d.)

References

- Bloomberg Associates. n.d. "Equitably Expanding Chicago's Tree Canopy." Bloomberg Associates. Accessed May 21, 2024. <https://associates.bloomberg.org/cities/chicago/equitably-expanding-chicagos-tree-canopy/story/what-tree-equity-new-tool-uchicago-data-scientists-helping-transform-neighborhood-health>
- Caney, Simon. 2005. "Cosmopolitan Justice, Responsibility, and Global Climate Change." *Leiden Journal of International Law* 18 (4): 747–75.
- Chaudhury, Pourabi, and Debanjan Banerjee. 2020. "Recovering with Nature': A Review of Ecotherapy and Implications for the COVID-19 Pandemic." *Frontiers in Public Health* 8.
- City of Chicago. n.d.a. "Chicago Recovery Plan Home." City of Chicago. Accessed May 20, 2024. <https://www.chicago.gov/city/en/sites/chicago-recovery-plan/home.html>.
- City of Chicago. n.d.b. "Our Roots Chicago." City of Chicago. Accessed May 20, 2024. <https://www.chicago.gov/city/en/sites/our-roots-chicago/home.html>.
- City of Chicago. n.d.c. "Tree Canopy Equity Expansion." Chicago Recovery Plan. Accessed May 20, 2024. <https://chirecoveryplan.com/programs/tree-canopy-equity-expansion/>.
- City of Chicago. 2021. "City Council Approves Revised Affordable Requirements Ordinance." Mayor's Office for People with Disabilities. <https://www.chicago.gov/city/en/depts/mopd/provdrs/hous/news/2021/march/mayor-lightfoot-and-department-of-housing-introduce-revised-affo.html>.
- City of Chicago. 2023. "Chicago Urban Forest Management Plan." Streets and Sanitation. Accessed May 19, 2024. <https://www.chicago.gov/city/en/depts/streets/provdrs/forestry/svcs/chicago-urban-forest-management-plan-.html>.
- Clarke, Mysha, Stephanie Cadaval, Charles Wallace, Elsa Anderson, Monika Egerer, Lillian Dinkins, and Ricardo Platero. 2023. "Factors That Enhance or Hinder Social Cohesion in Urban Greenspaces: A Literature Review." *Urban Forestry & Urban Greening* 84: 127936.
- Environmental Protection Agency. 2024. "Heat Island Effect." U.S. EPA. <https://www.epa.gov/heatislands>.
- Faber, Jacob William, and Jocelyn Pak Drummond. 2024. "Still Victimized in a Thousand Ways: Segregation as a Tool for Exploitation in the Twenty-First Century." *Annual Review of Sociology* 50.
- Forest Preserves of Cook County. 2022. "Our Ecosystems." Forest Preserves of Cook County. <https://fpdcc.com/nature/a-tour-of-our-ecosystems/>.
- Francis, Meredith. n.d. "The Story of Logan Square." WTTW. <https://www.wttw.com/chicago-by-1/neighborhoods/logan-square>.
- Gilligan, Carol. 1993. *In a Different Voice: Psychological Theory and Women's Development*. Cambridge, MA: Harvard University Press.
- Graf, Carly. 2019. "Chicago's Green Space: Inequitable for 100 Years." *Medill Reports Chicago*. <https://news.medill.northwestern.edu/chicago/friends-of-the-parks-alleges-chicago-green-spaces-still-map-racial-in-equality/>.
- Harris, Brandon, Dorothy Schmalz, Lincoln Larson, Mariela Fernandez, and Sarah Griffin. 2020. "Contested Spaces: Intimate Segregation and Environmental Gentrification on Chicago's 606 Trail." *City & Community* 19 (4): 933–62.
- Intergovernmental Panel on Climate Change. 2021. "Climate Change Widespread, Rapid, and Intensifying – IPCC." IPCC. <https://www.ipcc.ch/2021/08/09/ar6-wg1-20210809-pr/>.

- Intergovernmental Panel on Climate Change. n.d. "FAQ 1.3: Historical Overview of Climate Change Science." AR4 WGI Chapter 1. Accessed May 21, 2024. https://archive.ipcc.ch/publications_and_data/ar4/wg1/en/faq-1-3.html.
- Jelks, Na'Taki Osborne, Viniene Jennings, and Alessandro Rigolon. 2021. "Green Gentrification and Health: A Scoping Review." *International Journal of Environmental Research and Public Health* 18 (3): 907.
- Kim, Sage J., Caroline Kery, Jinghua An, James Rineer, Georgiy Bobashev, and Alicia K. Matthews. 2024. "Racial/Ethnic Disparities in Exposure to Neighborhood Violence and Lung Cancer Risk in Chicago." *Social Science & Medicine* 340: 116448.
- Lee, Jong-Won, Sang-Woo Lee, Hai Gyong Kim, Hyun-Kil Jo, and Se-Rin Park. 2023. "Green Space and Apartment Prices: Exploring the Effects of the Green Space Ratio and Visual Greenery." *Land* 12 (11): 2069.
- Liboiron, Max. 2021. *Pollution is Colonialism*. Durham: Duke University Press.
- Manca, Anna Rita. 2014. "Social Cohesion." In *Encyclopedia of Quality of Life and Well-Being Research*, edited by Alex C. Michalos, 6026–28. Dordrecht: Springer.
- Martinez, Deniss J, Cannon, Clare E. B., McInteruff, Alex, Alagona, Peter S., Pel-low, David N. 2023. "Back to the future: Indigenous relationality, kincentricity and the North American Model of wildlife management." *Environmental Science & Policy* 140: 202-7. <https://doi.org/10.1016/j.envsci.2022.12.010>.
- Matthews, Stephen N., Louis R. Iverson, Matthew P. Peters, and Anantha M. Prasad. 2018. "Assessing Potential Climate Change Pressures Across the Conterminous United States: Mapping Plant Hardiness Zones, Heat Zones, Growing Degree Days, and Cumulative Drought Severity Throughout This Century." RMAP-NRS-9. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northern Research Station. <https://doi.org/10.2737/NRS-RMAP-9>.
- NASA. n.d. "The Earth's Radiation Budget." NASA Science. Accessed May 19, 2024. https://science.nasa.gov/ems/13_radiation-budget/.
- National Park Service. 2022. "Carbon Storage by Urban Forests." National Park Service. <https://www.nps.gov/articles/000/uer-la-trees-carbon-storage.htm>.
- Parrella-Aureli, Ariel. 2024. "Anti-Gentrification Ordinance Protecting Homes Near The 606 And Pilsen Could Be Extended." Block Club Chicago, March 7, 2024. <https://blockclubchicago.org/2024/03/07/anti-gentrification-ordinance-protecting-homes-near-the-606-and-pilsen-could-be-extended/>.
- Qin, Yinghong. 2020. "Urban Flooding Mitigation Techniques: A Systematic Review and Future Studies." *Water* 12 (12): 3579. <https://doi.org/10.3390/w12123579>.
- Rahman, Mohammad A., Astrid Moser, Thomas Rötzer, and Stephan Pauleit. 2017. "Within-Canopy Temperature Differences and Cooling Ability of *Tilia cordata* Trees Grown in Urban Conditions." *Building and Environment*, 114: 118–28.
- Rigolon, Alessandro, Samuel J. Keith, Brandon Harris, Lauren E. Mullenbach, Lincoln R. Larson, and Jaclyn Rushing. 2020. "More Than 'Just Green Enough': Helping Park Professionals Achieve Equitable Greening and Limit Environmental Gentrification." *Journal of Park & Recreation Administration* 38, (3): 29-54.
- Rucks-Ahidiana, Zawadi. 2022. "Theorizing Gentrification as a Process of Racial Capitalism." *City & Community* 21 (3): 173–92.
- Summers, James K., and Deborah N. Vivian. 2018. "Ecotherapy—A Forgotten Ecosystem Service: A Review." *Frontiers in Psychology* 9: 1389.
- Turner-Skoff, Jessica B., and Nicole Cavender. 2019. "The Benefits of Trees for Livable and Sustainable Communities." *Plants, People, Planet* 1 (4): 323–35.

- United States Department of Agriculture. n.d.a. "How to Use the Maps." USDA Plant Hardiness Zone Map. Accessed May 19, 2024. <https://planthardiness.ars.usda.gov/pages/how-to-use-the-maps>.
- United States Department of Agriculture, Natural Resources Conservation Service. n.d.b. "A Conservation Plant Release by USDA NRCS Manhattan Plant Materials Center, Manhattan, Kansas." Plant Materials Program. Accessed May 21, 2024.
- United States Department of Agriculture. Office of Sustainability and Climate. 2021. Climate Change Pressures in the 21st Century. <https://storymaps.arcgis.com/stories/9ee0c-c0a070c409cbde0e3a1d87a487c>.
- Whyte, Kyle Powys. 2014. "Indigenous Women, Climate Change Impacts and Collective Action." *Hypatia: A Journal of Feminist Philosophy* 29 (3): 599–616.
- Whyte, Kyle Powys. 2017. "Indigenous Climate Change Studies: Indigenizing Futures, Decolonizing the Anthropocene." *English Language Notes* 55 (1): 153–62.
- Whyte, Kyle Powys. 2018. "Indigenous Science (Fiction) for the Anthropocene: Ancestral Dystopias and Fantasies of Climate Change Crises." *Environment and Planning E: Nature and Space* 1 (1-2): 224–42.
- Whyte, Kyle Powys. 2020. "Too Late for Indigenous Climate Justice: Ecological and Relational Tipping Points." *Wiley Interdisciplinary Reviews: Climate Change* 11 (1): e603.
- Whyte, Kyle Powys, Chris J. Cuomo, S. M. Gardiner, and A. Thompson. 2016. *Ethics of Caring in Environmental Ethics*. Oxford: Oxford University Press.
- Witynski, Max. 2021. "What Is Tree Equity? A New Tool from UChicago Data Scientists Is Helping to Transform Neighborhood Health." *UChicago News*. <https://news.uchicago.edu/story/what-tree-equity-new-tool-uchicago-data-scientists-helping-transform-neighborhood-health>.

Protecting the Digital Town Square

A Privacy-Preserving Framework To Combat AI-Powered Social Media Bots

Florin Schoenborn¹

Abstract

Breakthroughs with large language models (LLMs) have made social media bot networks much cheaper to operate and harder to detect. Using millions of dormant, inauthentic accounts, these networks can be used to amplify targeted narratives, flood platforms with inauthentic content, and scam users. As citizens become less certain about what is real online, they develop a deep distrust and engage less in online conversations, fueling polarization and undermining a core mechanism of democratic self-governance: the free exchange of ideas. Despite these risks, social media platforms (SMPs) have limited incentives to detect, disclose, or remove bot activity. Researchers face growing restrictions on data access, and policymakers lack information about the scale of AI-driven manipulation, even as bots improve rapidly. To restore transparency and curb bot activity, this memo proposes a “Trust, but Verify” framework. Under this approach, social media platforms are trusted with the responsibility to detect bot activity on their own terms, but the adequacy of their efforts is independently verified. To achieve this, rather than regulating content or access to social media, which could trigger First Amendment challenges, Congress should mandate that SMPs submit biannual “bot transparency reports” to the Federal Trade Commission (FTC) using its Section 6(b) investigatory authority. These reports would detail bot detection methodologies, estimated bot prevalence, and removal performance, subject to review by independent researchers. To ensure compliance, Congress should pass a new penalty structure: if SMPs repeatedly fail to detect or report bot activity, escalating fines should incentivize them to temporarily introduce user authentication during signup. Since user authentication hurts conversion rates and revenue, SMPs are then incentivized to quickly restore transparency and/or bot detection accuracy. To ensure that such temporary user authentication protects the privacy of citizens while effectively curbing bot activity, Congress should fund research and development (R&D) into a one-time, zero-knowledge proof user authentication framework implemented at the operating-system (OS) level (e.g., iOS and Android). Under this system, users verify their identity once when purchasing or setting up a device, and a privacy-preserving credential is stored locally on the OS-level. If an SMP activates temporary authentication due to an inability to detect bots, users simply present this stored credential—no additional identity disclosure is required. Once the platform’s bot detection performance improves and fines are lifted, the authentication is removed, restoring frictionless access. This approach avoids the legal, technical, and political failures of age-verification schemes, which require users to submit personally-identifiable data to third parties. Instead, it provides SMPs with a clearly defined, legally sound, and effective approach to curb bot activity without endangering user privacy.

Background: The Scale and Nature of the Bot Problem

Growing Use of AI-Powered Bots

The use of AI-powered bots has grown rapidly in recent years. In 2024, automated web traffic exceeded human web traffic for the first time in a decade, accounting for 51% of all traffic; 37% of this automated traffic qualified as coming from “bad bots,” defined as automated software with malicious or deceptive intent

(Imperva 2025). This surge is widely attributed to advances in LLMs, which have lowered the technological barrier and cost for large-scale bot networks. Since 2022, the cost of operating an AI-powered account on X (formerly Twitter) fell by a factor of four, with similar cost declines observed across major platforms (Bergmanis-Koráts and Haiduchyk 2024).

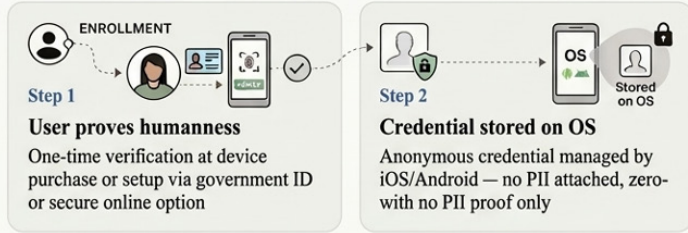
¹ Florin Schoenborn is an MPP student jointly at the McCourt School of Public Policy at Georgetown University and the Hertie School in Berlin, Germany

SUMMARY

1

Framework: anonymous credential

A one-time, zero-knowledge proof credential is created at device setup and stored at the OS level, with no personally identifiable information tied to it.



2

Use: trust, but verify

SMPs must transparently report bot detection capability. If detection fails, escalating fines push platforms toward temporary user authentication — which the credential makes frictionless.

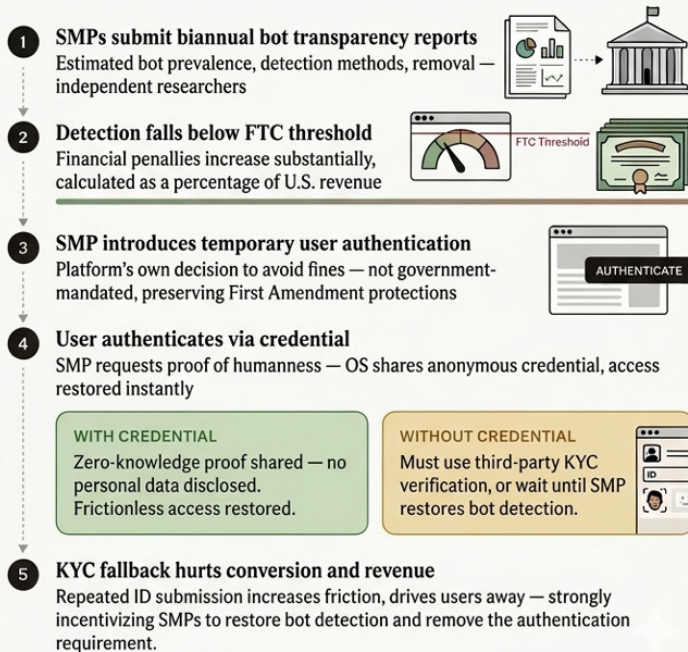


Figure 1: Summary of user authentication framework and “trust, but verify” rules for bot detection

Humans Can No Longer Detect AI-Powered Bots

At the same time, humans are increasingly unable to distinguish AI-driven accounts from genuine ones. In a controlled 2025 study, participants judged GPT-4 to be human 63.7% of the time in conversational settings (Jones et al. 2025). A separate 2024 study found that human detection of AI-generated content across text, images, audio, and video performed no better than chance (Cooke et al. 2024).

Bot Detection and Removal by SMPs is Inadequate

Bot network operators are using physical phone farms, SIM banks, mobile proxies, IP geolocation spoofing, and device fingerprinting to avoid detection (Schwartzman 2025). As a result, SMP bot removal rates are falling. A NATO red-teaming exercise revealed that in 2024, with the partial exception of X, which removed approximately 50% of bot accounts, major platforms such as YouTube, TikTok, Instagram, and Facebook removed only 2 to 25% of bot accounts (Bergmanis-Korāts and Haiduchyk 2024). This difference between platforms showcases the ongoing arms race between bots and detection algorithms. The emergence of venture capital-backed “AI influencer” startups further illustrates how effective and widespread

social media bots have become. For example, DoubleSpeed AI, a startup explicitly marketing AI-driven bot networks designed to replace human influencers, has received funding from Andreessen Horowitz (Wilkins 2025).

SMPs are Incentivized Against Bot Transparency

SMPs derive revenue from impressions and engagement, regardless of whether an account is human- or bot-powered. Any disclosure of widespread bot activity could undermine advertiser confidence and user trust, creating strong incentives to minimize or obscure the problem (Wilkins 2025). As a result, SMPs have curtailed researcher access to data, making independent measurement of bot prevalence increasingly difficult (Brennen, Sanderson, and de la Puerta 2025).

The Abuse Potential Mandates Urgent Action

While some studies emphasize that persistent data gaps make research difficult, the abuse potential for AI-powered social media bots clearly warrants action. For example, the FBI warns of a surge in AI-generated kidnapping scams, where parents are presented artificially-generated images of their children in captivity (Brennen, Sanderson, and de la Puerta 2025). Social media bots have a measurable influence on sentiment dynamics, which has been exploited by foreign governments (Luo et al. 2023). Furthermore, the presence of human-like bots has degraded trust in online discourse, as reflected in the sharp rise in social media users accusing others of being “bots” over the last ten years (Assenmacher, Fröhling, and Wagner 2024). Because advances in technology will further reduce the effectiveness of bot detection, social media platforms should be legally required to disclose current detection performance and bot prevalence, and to follow clear regulatory guidance on how to address future generations of automated accounts.

Lessons from Existing Regulatory Approaches

Bot Labeling and Disclosure Requirements Failed

Existing regulatory approaches in the United States and Europe have focused on bot labeling, disclosure, and prohibition of malicious automated behavior on SMPs and by AI companies. Examples include Article 50(1) of the EU AI Act, the California Bot Act (SB 1001), Utah’s Artificial Intelligence Consumer Protection Amendments, and recent FTC rules restricting the purchase of fake engagement (European Union 2024; California 2018; Utah 2025; Federal Trade Commission 2024).

While these measures established important norms—principally, the expectation that automated accounts should be disclosed to users, that platforms bear responsibility for policing inauthentic behavior, and that purchasing fake engagement constitutes a deceptive trade practice—they were insufficient in practice. Labeling and disclosure requirements placed the burden of compliance on malicious bot operators, while broad prohibitions incentivized SMPs to delete any account suspected of being a bot, regardless of certainty, leading to the removal of many legitimate accounts and chilling free speech (McSherry 2018). Enforcement was further complicated by the frequent location of bot operators outside of US jurisdiction (Williams 2019).

Only California’s relatively narrow definition of prohibited bots—requiring “intent to mislead the other person [...] in order to incentivize a purchase or influence a vote”—has reduced the risk of legitimate accounts being removed. However, it also made it difficult to remove any bot account, as proving intent was difficult (Radivojevic et al. 2024). The key lesson from these efforts is that effective bot regulation must primarily target platforms’ detection capabilities rather than bot operator disclosure or inaccurate reporting features.

Age Verification Schemes Caused Backlash and Were Mostly Ineffective

Recent age-verification regimes offer additional lessons for user verification frameworks, which could also be used to curb inauthentic accounts. Laws such as the UK Online Safety Act, Australia’s Online Safety

Amendment, Utah’s Social Media Regulation Act (SB 152), and California’s Digital Age Assurance Act demonstrated the pitfalls of user authentication for social media access (United Kingdom 2023; Australia 2024; Utah 2023; California 2025).

Many of these regimes relied on third-party “know your customer” providers that required users to repeatedly submit government IDs and biometric data to prove their age, triggering widespread public backlash and significant privacy risks (Prescod 2024). Studies suggest that such schemes also encouraged circumvention through VPN use and imposed disproportionate burdens on lawful users (Lang et al. 2025). Moreover, fear-driven overcompliance has led to much lawful and child-appropriate content being hidden behind age verification screens (Castro and Ambrose 2025). Finally, in the United States, age-verification laws have been successfully challenged on First Amendment grounds (NetChoice, LLC v. Griffin 2025).

A notable exception is California’s Digital Age Assurance Act (CDAA), for which parents verify a child’s age just once when setting up their smartphone or tablet. This information is then stored by the operating system (iOS, Android) and shared with apps and websites without personally identifiable information. Because no repeated verification or identity disclosure is required, this model has attracted support from major SMPs and hardware providers such as Apple and Google (Wicks 2025). Given its effectiveness, simplicity, and privacy-preserving nature, this memo adopts the CDAA’s approach of one-time, OS-level user verification to curb AI-powered bot activity.

Policy Recommendations: A “Trust, but Verify” Framework for Bot Detection

1. Mandate SMPs to Accurately and Transparently Report Their Bot Detection Capabilities

Congress should direct the FTC to use its existing Section 6(b) authority to require SMPs with more than 50 million US users to submit biannual “bot transparency reports” (Federal Trade Commission 2025). These reports should include:

- Estimated bot prevalence on the platform;
- Bot detection methodologies and perfor-

mance metrics; and

- Removal and mitigation rates over time.

To ensure accountability, platforms must grant independent “vetted researchers,” modeled on the EU Digital Services Act framework, access to relevant data and methodologies for verification. Because Section 6(b) authority is limited to investigations, Congress must also legislate new financial penalties for noncompliance, refusal to provide data access, or material misrepresentation, triggering fines for SMPs calculated as a percentage of US revenue.

2. Use Fines to Incentivize Temporary User Authentication if Bot Detection Fails

If a platform’s reported bot prevalence or detection performance hits an FTC-defined critical threshold, financial penalties should increase substantially. At this stage, platforms would face strong incentives to adopt temporary user authentication measures to restore compliance, as this will effectively curb the creation and use of AI-powered bot accounts not verifiably tied to a human user. Importantly, the government would not mandate the use of user authentication. The decision to restrict access would remain with private platforms, preventing First Amendment challenges while ensuring that inaction is financially painful.

The FTC’s 2019 settlement with Facebook (now Meta) offers a useful precedent for calibrating such penalties: the \$5 billion fine—then the largest ever imposed by the FTC—represented approximately one month of Facebook’s revenue and was large enough to compel the company to restructure its entire privacy governance, including the creation of an independent board-level privacy committee and mandatory CEO compliance certifications, yet it did not threaten the platform’s viability (Federal Trade Commission 2019). A similar revenue-based penalty structure, scaled to bot detection performance, could provide the strong compliance incentive this framework requires without crossing the state-action threshold.

The user authentication is designed to be temporary: once a platform’s bot detection performance returns to an acceptable level, the escalated fines are removed, and with them the incentive to maintain the authentication requirement. Platforms are strongly motivated to make verification temporary rather than permanent because requiring user

authentication as a permanent part of the sign-up process would increase friction and reduce conversion rates, harming revenue.

3. Ensure SMPs Use Open-Source, Privacy-Preserving User Authentication Framework

To prepare for scenarios in which bots become hard to detect, and where SMPs are thus incentivized to temporarily introduce user authentication, Congress should fund Defense Advanced Research Projects Agency (DARPA) R&D for a one-time, zero-knowledge proof, and open-source user authentication framework.

Such a framework should be based on California’s Digital Age Assurance Act and the proposals for “personhood credentials” by OpenAI researchers, adhering to these principles (Adler et al. 2024):

- One-time authentication: Verification occurs once per device, unless device ownership changes.
- Operating-system level: Credentials are stored and managed at the operating-system level (iOS or Android), rather than by individual platforms or websites.
- Zero-knowledge proof: No personally identifiable information is shared during authentication; only the fact that a unique human user is associated with the device.
- Equitable access: Multiple pathways to authentication must be available, including at device purchase, through government-run services, or via secure online options.

By January 2027, DARPA should publish a technical and governance framework for such a system, in coordination with hardware providers (e.g., Apple, Google, Samsung) and SMPs. Congress should then require the FTC to ensure that any user authentication process adopted by SMPs must be based on this framework to protect user privacy.

Addressing Counterarguments

Any Government Mandated User Authentication Infringes on the First Amendment

Because the government does not regulate the content of users’ speech or mandate user authentication, but instead imposes transparency obligations and financial penalties

on platforms to influence their terms of service, it is less likely to face First Amendment challenges. The First Amendment does not prevent private companies from restricting access to their platforms for users who refuse to authenticate.

Citizens Will See This as Government Overreach

To ensure citizens will not view user authentication as arbitrary or government-mandated, SMPs are required to share the findings of their biannual bot reports with their users. If the prevalence of AI-powered bots becomes too high and requires intervention to avoid fines, SMPs can frame user authentication as a temporary-but-necessary measure to restore a better user experience. In the end, users do not like the presence of bots, and they will welcome attempts at restoring a more “human” social media experience. Finally, SMPs are incentivized to highlight the one-time, fully anonymous nature of authentication, since users might otherwise stop using their platform altogether due to privacy concerns.

Using Fines to Coerce SMPs to Adopt User Authentication Risks Creating “State Action”

This concern is valid. Financial penalties enacted by Congress cannot be so high that SMPs have no choice but to adopt user authentication in order to remain in business. Such fines would constitute state action, meaning the government would be using its regulatory power to turn a private entity into an actor carrying out the will of the state (US Constitution Annotated n.d.). In that case, SMPs would no longer be acting in a private capacity, and First Amendment challenges could be brought directly against them. To avoid this result, SMPs must retain the ability to reject user authentication, even if doing so requires substantial fines for failing to address bot activity. Financial penalties can be significant, but they cannot threaten the survival of SMPs.

Conclusion

AI-powered social media bots pose a high potential for abuse and threaten the destruction of online discourse, with current regulatory approaches proving inadequate to stop their proliferation. This policy memo argues for a pragmatic “Trust, but Verify” solution: mandate

biannual bot transparency reports for SMPs, enforce civil penalties for noncompliance, and invest in R&D for privacy-preserving user authentication frameworks as a last resort. This strategy equips policymakers with better data to understand the scale of the bot problem, while balancing SMP accountability with privacy protections. Finally, by investing in user authentication frameworks for a future in which bots are indistinguishable from humans, Congress can help ensure that social media remains a space reserved for humans, not AI-powered bots.

References

- Adler, S., Z. Hitzig, S. Jain, C. Brewer, W. Chang, R. DiResta, E. Lazzarin, et al. 2024. "Artificial Intelligence and the Value of Privacy-Preserving Tools to Distinguish Who Is Real Online." arXiv, August 2024. <https://arxiv.org/abs/2408.07892>.
- Assenmacher, D., L. Fröhling, and C. Wagner. 2024. "You Are a Bot! Studying the Development of Bot Accusations on Twitter." *Proceedings of the International AAAI Conference on Web and Social Media* 18 (1): 113-125. <https://doi.org/10.1609/icwsm.v18i1.31301>.
- Australia. 2024. Online Safety Amendment (Social Media Minimum Age) Bill 2024. Parliament no. 47 (2024). https://www.aph.gov.au/Parliamentary_Business/Bills_Legislation/Bills_Search_Results/Result?bid=r7284.
- Bergmanis-Koräts, G., and T. Haiduchyk. 2024. *Social Media Manipulation for Sale: Experiment on Platform Capabilities to Detect and Counter Inauthentic Social Media Engagement*. Riga: NATO Strategic Communications Centre of Excellence. <https://stratcomcoe.org/publications/social-media-manipulation-for-sale-experiment-on-platform-capabilities-to-detect-and-counter-inauthentic-social-media-engagement/311>.
- Brennen, S. B., Z. Sanderson, and C. de la Puerta. 2025. "When It Comes to Understanding AI's Impact on Elections, We're Still Working in the Dark." Brookings Institution, March 4, 2025. <https://www.brookings.edu/articles/when-it-comes-to-understanding-ais-impact-on-elections-were-still-working-in-the-dark/>.
- California. 2018. Senate Bill 1001 (2017-2018 Regular Session): Bots: Disclosure, chaptered September 28, 2018. Digital Democracy, CalMatters. https://calmatters.digitaldemocracy.org/bills/ca_20172018sb1001.
- California. 2025. Assembly Bill 1043 (2025-2026 Regular Session): Age Verification Signals: Software Applications and Online Services. Chap. 675, Statutes of 2025, added Title 1.81.9 to Part 4 of Division 3 of the Civil Code (approved October 13, 2025). https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202520260AB1043.
- Castro, D., and A. Ambrose. 2025. "The UK's Online Safety Act's Predictable Consequences Are a Cautionary Tale for America." Information Technology and Innovation Foundation, September 3, 2025. <https://itif.org/publications/2025/09/03/the-uks-online-safety-acts-predictable-consequences-are-cautionary-tale-for-the-us/>.
- Cooke, D., A. Edwards, S. Barkoff, and K. Kelly. 2024. "As Good as a Coin Toss: Human Detection of AI-Generated Images, Videos, Audio, and Audiovisual Stimuli." arXiv. <https://arxiv.org/abs/2403.16760>.
- European Union. 2024. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 12 June 2024 on Artificial Intelligence (Artificial Intelligence Act). Article 50, "Transparency Obligations for Providers and Deployers of Certain AI Systems." <https://artificialintelligenceact.eu/article/50/>.
- Federal Trade Commission. 2019. "FTC Imposes \$5 Billion Penalty & Sweeping New Privacy Restrictions on Facebook." Press release, July 24, 2019. <https://www.ftc.gov/news-events/news/press-releases/2019/07/ftc-imposes-5-billion-penalty-sweeping-new-privacy-restrictions-facebook>.
- Federal Trade Commission. 2024. "Federal Trade Commission Announces Final Rule Banning Fake Reviews and Testimonials." Press release, August 14, 2024. <https://www.ftc.gov/news-events/news/press-releases/2024/08/federal-trade-commission-announces-final-rule-banning-fake-reviews-testimonials>.

- Federal Trade Commission. 2025. Enforcement Authority. Revised July 2025. <https://www.ftc.gov/about-ftc/mission/enforcement-authority>.
- Imperva. 2025. 2025 Bad Bot Report. <https://www.imperva.com/resources/resource-library/reports/2025-bad-bot-report/>.
- Jones, C. R., I. Rathi, S. Taylor, and B. K. Bergen. 2025. "People Cannot Distinguish GPT-4 from a Human in a Turing Test." In Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency, 1615-1639. <https://arxiv.org/abs/2405.08007>.
- Lang, D., B. Listyg, B. V. Ross, A. V. Musquera, and Z. Sanderson. 2025. "Do Age-Verification Bills Change Search Behavior? A Pre-Registered Synthetic Control Multiverse." Open Science Framework. <https://osf.io/z83ev>.
- Luo, H., X. Meng, Y. Zhao, and M. Cai. 2023. "Rise of Social Bots: The Impact of Social Bots on Public Opinion Dynamics in Public Health Emergencies from an Information Ecology Perspective." *Telematics and Informatics* 85: 102051. <https://doi.org/10.1016/j.tele.2023.102051>.
- McSherry, C. 2018. "Platform Censorship: Lessons from the Copyright Wars." Electronic Frontier Foundation, September 26, 2018. <https://www.eff.org/deeplinks/2018/09/platform-censorship-lessons-copy-right-wars>.
- NetChoice, LLC v. Griffin. 2025. No. 5:23-CV-05105 (W.D. Ark. Mar. 31, 2025). https://netchoice.org/wp-content/uploads/2025/04/Court-Permanently-Halts-Arkansas-Age-Verification-Law_NetChoice-v-Griffin_Mar-31-2025.pdf.
- Prescod, M. 2024. "Why Online Age Verification Will Give Us the Worst of Both Worlds." Institute of Economic Affairs, October 2, 2024. <https://iea.org.uk/why-online-age-verification-will-give-us-the-worst-of-both-worlds/>.
- Radivojevic, K., C. McAleer, C. Conley, C. Kennedy, and P. Brenner. 2024. "Social Media Bot Policies: Evaluating Passive and Active Enforcement." arXiv. <https://doi.org/10.48550/arXiv.2409.18931>.
- Schwartzman, E. 2025. "Bot Farms Invade Social Media to Hijack Popular Sentiment." Fast Company, April 24, 2025. <https://www.fastcompany.com/91321143/bot-farms-social-media-manipulation>.
- United Kingdom. 2023. Online Safety Act 2023, 2023 c. 50. UK Public General Acts. <https://www.legislation.gov.uk/ukpga/2023/50/>.
- U.S. Constitution Annotated. n.d. "Amdt1.7.2.4 State Action Doctrine and Free Speech." Library of Congress. https://constitution.congress.gov/browse/essay/amdt1-7-2-4/ALDE_00013541/.
- Utah. 2023. Senate Bill 152 (2023 General Session): Social Media Regulation Amendments, signed March 23, 2023 (effective May 3, 2023). <https://le.utah.gov/~2023/bills/static/SB0152.html>.
- Utah. 2025. Senate Bill 226 (2025 General Session): Artificial Intelligence Consumer Protection Amendments, enacted March 27, 2025. <https://le.utah.gov/~2025/bills/static/SB0226.html>.
- Wicks, B. 2025. "Google, Meta among Tech Leaders and Child Advocates Voicing Support for Wicks' Digital Age Assurance Act." California State Assembly, September 9, 2025. <https://a14.asmdc.org/press-releases/20250909-google-meta-among-tech-leaders-and-child-advocates-voicing-support-wicks>.
- Wilkins, J. 2025. "AI 'Phone Farm' Startup Gets Funding from Marc Andreessen to Flood Social Media with Spam." Futurism, October 27, 2025. <https://futurism.com/artificial-intelligence/doublespeed-ai-phone-farm>.
- Williams, J. L. 2019. "Cavalier Bot Regulation and the First Amendment's Threat Model." Knight First Amendment Institute, August 21, 2019. <https://knightcolumbia.org/content/cavalier-bot-regulation-and-the-first-amendments-threat-model>.

Addressing Insulin Affordability

Policy Pathways to Expand Access, Reduce Costs, and Improve Health

Rachel Williams¹

Abstract

Insulin affordability is a persistent public health concern in the United States. In 2021, approximately 38.4 million Americans were living with diabetes, many of whom rely on insulin as a medically essential treatment. Demand for insulin is highly inelastic, yet United States insulin prices remain substantially higher than those in comparable countries. As a result, many patients ration their medication, leading to adverse health outcomes. The principal contributing factor to the high prices for insulin in the United States is the oligopolistic structure of the insulin production market, which is dominated by three major pharmaceutical firms. Recent federal policies, including the Affordable Care Act (2010), the Inflation Reduction Act (2022), and the Affordable Insulin Now Act (2023), reduce insulin costs for insured individuals through expanded coverage and caps on monthly cost-sharing. However, these policies do not directly address affordability for uninsured individuals, who continue to face high out-of-pocket costs and limited access to insulin. This paper proposes a policy alternative that reduces barriers to competition by subsidizing entry into the insulin production market. The subsidy in this approach is meant to offset fixed entry costs such as regulatory approval, manufacturing infrastructure, and research and development. It uses a cost-benefit analysis framework to compare this targeted subsidy approach with two common approaches: maintaining the status quo, and broad price-cap policies. The analysis suggests that subsidizing market entry has the potential to increase competition and lower insulin prices over time, while addressing affordability concerns that remain unresolved under existing policy interventions.

The Rising Cost of Insulin: Impact, Causes, and Current Policy Responses

Insulin is essential for life. Without it, glucose can build up in the blood, which prevents all organs from receiving necessary energy, and ultimately leads to organ failure (Johns Hopkins Medicine n.d.). Diabetes is a chronic disease that occurs when the pancreas does not produce enough insulin or when the body cannot effectively use the insulin it produces (World Health Organization 2024). According to the United States Centers for Disease Control and Prevention (2024), 38.4 million people in the United States had diabetes in 2021 and the number is similar today. Among this population, approximately 1.56 million adults were uninsured in 2023, highlighting a major gap in access to affordable insulin (Casagrande et al. 2023).

Although prescribed insulin is essential for people with diabetes (Prasad 2019), the cost of insulin in the United States leads many patients to ration it. This means they stretch

their prescription, do not fill it, or use less than prescribed because of cost. A study conducted at the Yale Diabetes Center found that 1 in 4 patients in the study rationed their insulin (Crawford 2023). Using less insulin than prescribed can cause diabetic ketoacidosis, kidney failure, amputations, blindness, heart attacks, strokes, or other complications (Crawford 2023).

Many Americans risk these complications in order to reduce spending on insulin. Rationing is especially prevalent for individuals in middle or lower-income socioeconomic groups. Insulin prices in the United States increased by about 1000% from 1923 to 2019 (Prasad 2019). For example, one vial of insulin lispro cost \$21 in 1999, but \$332 in 2019 (Rajkumar 2020). While these figures are in nominal terms, even after adjusting for inflation, insulin prices have increased substantially over this period. Many patients require 12–36 vials of insulin annually (Chesak 2022). Given the average U.S. income of about \$67,000, this represents a substantial financial burden, making insulin virtually unaffordable for many Americans (Social Security Administration 2023).

¹ Rachel Williams is an undergraduate student at the University of Wisconsin-Madison

The high price of insulin in the United States is influenced by several factors, including a small number of producers, high barriers to entry in the pharmaceutical industry, and the highly inelastic demand for insulin. The insulin market is primarily shaped by 3 major pharmaceutical companies—Eli Lilly, Sanofi, and Novo Nordisk (Crawford 2023). Due to the complex nature of the pharmaceutical industry and limited competition, these manufacturers have historically adjusted prices in response to limited competition among a small number of dominant firms. Insulin plays an essential role in diabetes management and has highly inelastic demand. As a result, firms' incentives to sustain innovation, research, profit, and business operations, leads to price increases (Rajkumar 2020).

Recent developments in the pharmaceutical market suggest that increasing competition through new entrants may be a viable strategy for lowering drug prices (DrugPatentWatch 2025). For example, nonprofit pharmaceutical manufacturer Civica Rx was created to increase competition and stabilize supply for essential medications, including insulin (American Hospital Association 2023). Similarly, the Mark Cuban Cost Plus Drug Company has attempted to reduce pharmaceutical prices by bypassing traditional pharmaceutical distribution channels and offering medications with transparent pricing (RescueMeds 2025). These models demonstrate that increased competition can reduce drug prices. For example, Civica Rx has committed to offering certain generic medications at significantly lower and more stable prices, while the Mark Cuban Cost Plus Drug Company uses transparent pricing—typically a 15% markup plus a fixed fee—to offer drugs below traditional market rates (Nerandru et al. 2023). These examples illustrate that alternative production and distribution models may help reduce costs and improve access, supporting policies aimed at lowering barriers to market entry.

The United States government began implementing regulations related to insulin pricing and transparency in 2010. According to Myerson and Laiteerapong (2016), the 2010 Affordable Care Act (ACA) required all health insurance plans to provide at least some coverage for diabetes care without lifetime and annual caps. The ACA also prohibits health

insurance plans from considering diabetes status when calculating premiums. Insurers may no longer use diabetes status to decide whether to accept an applicant. The ACA vastly improved Americans' access to healthcare and diabetes care, but it does not cover the uninsured population. After the ACA went into effect, people who are uninsured pay between approximately \$144 and \$404 more in out-of-pocket healthcare costs per year than people with insurance (Lin et al. 2023).

The Biden Administration passed the Inflation Reduction Act in 2022 and the Affordable Insulin Now Act in 2023, both of which build on the ACA. The Inflation Reduction Act caps monthly co-payments on prescription costs, such as insulin, for certain Medicare enrollees at \$35 (Sayeed et al. 2023). Generally, to qualify for Medicare one must be 65 or older, have a disability, End-Stage Renal Disease, or Amyotrophic Lateral Sclerosis (U.S. Department of Health and Human Services n.d.). Thus, to make insulin further accessible, the Administration passed the Affordable Insulin Now Act in 2023 which reduces cost-sharing for insulin under private health insurance to either \$35 or 25% of a plan's negotiated price a month, whichever is less, beginning in 2024 (118th Congress 2024).

However, these acts only reduce insulin costs for people with insurance, leaving people without insurance to pay high out-of-pocket costs. (Lin et al. 2023). In 2024, 65% of Americans reported living paycheck to paycheck, with income largely covering only basic needs such as housing and food (ACA International 2024). Therefore, the cost of health insurance is prohibitive for many households. For example, individuals without employer-sponsored coverage, Medicare, Medicaid, or private insurance pay roughly \$477 per month, or about \$5,724 per year, for healthcare (Health Markets 2024). In 2023, 131.6 million Americans earned less than \$50,000 per year (Tierney 2024), and individuals in this income group may be required to spend 10% or more of their annual income on insurance. This level of spending is financially unsustainable for many households. As a result, many individuals remain uninsured.

Current insulin affordability policy is unaccommodating for the uninsured population, often assuming they can afford out-

of-pocket insulin costs. However, uninsured individuals paid about \$403 more out of pocket after the Affordable Care Act than insured patients (Lin et al. 2023), even though many cannot afford insurance in the first place (ACA International 2024). Although recent government interventions, such as the Affordable Care Act and the Inflation Reduction Act, have attempted to improve insulin affordability, more work is needed to ensure all diabetics have adequate access to insulin.

Primary Stakeholders

1. Diabetics. Access to insulin strongly impacts the health and quality of life of a diabetic, thus policies aimed at improving insulin affordability must consider the quality of life of diabetics and its impact on society. Diabetics need to be able to afford and receive enough insulin to manage their health. The way the United States distributes insulin and the inability of individuals to afford insulin prescriptions has caused many diabetics to suffer health consequences. Diabetes has been the eighth leading cause of death in the U.S. since 2020 (Centers for Disease Control and Prevention 2020) and 101,209 people in the U.S. died from diabetes in 2022 (Centers for Disease Control and Prevention 2022). Thus, it is crucial to consider how the current climate of insulin affordability greatly impacts the health of those with diabetes when developing policy solutions.
2. Pharmaceutical companies. Provided pharmaceutical companies control the supply of insulin, policy alternatives must ensure that insulin remains accessible to patients while preserving incentives for continued production. Before being regulated by government policies, Eli Lilly, Sanofi, and Novo Nordisk were greatly increasing insulin prices in lockstep given their mutual goal of maximizing supply, demand, and profit (Crawford 2023). These companies are interested in generating profits, maintaining their position in the market, and retaining control over pricing. Potentially adjusting pricing or offering alternative treatment options strongly affects these companies' profitability, operations, and public image.

3. Taxpayers. Because a large portion of government spending is funded by taxpayer revenue (Weilk and Sneiderman 2024), policy alternatives must be cost-effective and broadly supported by taxpayers. Taxpayers desire their funds to use taxpayer funds efficiently, increase access to health insurance, consider ensuring affordable access to healthcare for citizens, and maintain the integrity of public health programs like Medicare (FamiliesUSA 2025). At the same time, they seek to pay as little as possible for both taxes and health services (Navigator Research 2025). Thus, a cost-benefit analysis for potential solutions must include government use of taxpayer dollars.

The Role of Policy Intervention

The two main factors that affect insulin affordability and accessibility are lack of competition and inequity in insurance access.

1. Lack of competition. Given the inelastic demand and oligopolistic nature of the insulin market, insulin suppliers are able to increase prices or withhold inventory with little shift in demand. Since there are only 3 primary insulin producing pharmaceutical companies, they are able to control the market in order to sustain innovation, research, continue business operations, and keep pace with the other producers (Rajkumar 2020). Increasing prices have negatively impacted the health of diabetics (Crawford 2023), so policy intervention is necessary to regulate competition and pricing. The government is the only entity powerful enough to regulate the market in a way that creates an equitable price for diabetics while allowing businesses to operate profitably.
2. The equity issue caused by not addressing uninsured diabetics. Current government policy fails to help uninsured diabetics gain access to affordable insulin, which demonstrates the lack of equity in the insulin market. A substantial number of adults with diabetes remain uninsured (Casagrande et al. 2023). These individuals face significantly higher out-of-pocket costs than insured patients, often paying nearly double for care (Lin et al. 2023). Overall, policy intervention is necessary as it can be deemed unreasonable to only cap the cost of insulin for those with

insurance and assume that diabetics can afford the costs of insulin without support.

The Federal Government, compared to private actors, is best suited to solve these market failures because of its regulatory power and access to funding and other resources.

Criteria for Evaluating Policy Alternatives

1. **Cost-Effectiveness:** The goal of this criteria is to measure the costs and benefits of policy alternatives in terms of financial efficiency. Provided these policy alternatives are funded by taxpayers, they should be evaluated based on their financial costs as well as their broader economic and social impacts.
2. **Equity:** The goal of this criteria is to measure the reduction of disparities between insured and uninsured populations in access to insulin. Policy alternatives must consider that the uninsured population is likely unable to afford health insurance or out-of-pocket insulin costs.
3. **Feasibility:** The goal of this criteria is to measure the likelihood that the United States will adopt the alternative policy. In order for the policy alternative to be feasible, it must have long-term sustainability, appropriate scalability, and adequate stakeholder support, administrative capacity, and funding.

Alternative Courses of Action

1. **Maintain Current Policies (Status-Quo).** The Senate Finance Committee would continue to fund the ACA, the 2022 Inflation Reduction Act, and the 2023 Affordable Insulin Now Act. The government currently spends approximately 3% of taxpayer revenue, or \$15.68 billion, out of the around \$4.9 trillion of annual taxpayer revenue on insulin (Schrier et al. 2024; Fiscal Data 2024) and would continue to do so if it maintained the status-quo.
2. **Capping all insulin costs at \$35 regardless of insurance status.** The government would use tax revenue to cap out-of-pocket insulin costs for all diabetics at \$35. The estimated cost of this policy can be calculated using several parameters. There are approximately 1.56 million uninsured diabetics in the United States (Casagrande et al. 2023). The

average market price of a vial of insulin is about \$400 (Lin et al. 2023), and many patients require about 25 vials per year (Social Security Administration, 2023; Chesak, 2022). Under a \$35 price cap, the government would need to subsidize approximately \$365 per vial to cover the difference between the market price and the capped price. Multiplying the subsidy per vial by 25 vials per patient and 1.56 million uninsured individuals results in an estimated annual program cost of approximately \$14.2 billion.

3. **Subsidize entry costs to the market.** The government would lower entry costs to the insulin supplier market by directly subsidizing entry. These subsidies would cover expenses like manufacturing facilities, regulatory approval processes, research and development, and initial production setup to encourage more companies to become insulin producers. The proposed subsidy is not intended to fully finance insulin production. Instead, it is designed to offset a portion of the fixed costs associated with market entry, such as regulatory approval processes, facility upgrades, and initial production setup. Economic theory suggests that lowering fixed costs of entry increases the likelihood of new firms entering a market, thereby increasing competition (Buomal and Willig 1981). By lowering these entry barriers, the subsidy increases the likelihood that additional firms will enter the insulin market.

First, new producers would need to receive approval from the U.S. Food and Drug Administration to ensure they follow insulin production rules based on current regulations (U.S. Food and Drug Administration n.d.). Then, they would receive \$200,000 to help get them started and the opportunity to apply for \$100,000 for other start-up costs in the first 5 years of insulin production with a cap at \$500,000 spent on each new producer. The program would require approximately \$30 million in taxpayer funding to support entry by roughly 20 new insulin producers, a scale chosen to balance fiscal feasibility with meaningful increases in competition. Although the cost of developing and manufacturing biosimilar pharmaceuticals can exceed \$100 million due to research, regulatory approval, and manufacturing requirements (Blackstone &

Joseph 2013), the proposed subsidy is not intended to fully cover these costs. Instead, it is designed to offset a portion of the fixed entry costs and reduce financial risk, thereby encouraging firms that are already considering entry to participate in the market. Even modest reductions in fixed costs can significantly affect firm entry decisions in high-barrier markets. For example, research on sunk costs shows that even small reductions in entry costs can have large, persistent effects on firm entry and exit, particularly in industries with high upfront investments (Schmalensee 2004).

As a result, as additional firms enter the market, competitive pressure would make it more difficult for existing manufacturers to sustain high oligopolistic prices. Lower wholesale insulin prices would eventually translate into lower retail prices, which would directly reduce out-of-pocket costs for uninsured individuals who must pay the full market price for insulin. Since biosimilar drug development and regulatory approval typically takes years, the effects of increased market entry on insulin prices would likely occur gradually rather than immediately. Once new competitors enter

Table 1: Evaluation of Alternatives based on Criteria

Alternatives	Criteria		
	<u>Cost effectiveness:</u> Likelihood of balance between financial costs and societal and economic benefits	<u>Equity:</u> Likelihood of reduced disparities in insulin access for insured and uninsured people	<u>Feasibility:</u> Likelihood of successful implementation, sustainability, and stakeholder acceptance
Status Quo	Fair. Maintaining the status-quo caps insulin at \$35 for the insured. This makes insulin significantly more affordable for ~94% of diabetics in the US. However, it does not make insulin affordable for the 1.56 million diabetics who are uninsured. Society would have to pay for long run health costs associated with people who ration or choose to not take medication.	Fairly poor. Current policies provide support for most diabetics, but do not provide support for uninsured diabetics. Insurance or out-of-pocket insulin costs are not accessible to everyone, especially for those living paycheck to paycheck.	Good. This has already been implemented and can be sustained for a long period of time due to existing legislation budgeting. However, diabetics likely want to make insulin more affordable. Pharmaceutical companies continue to face pricing pressure from oligopolistic competition and balancing costs and benefits. Taxpayers who are interested in the overall health of society may not be content not supporting the uninsured.
Capping all insulin costs at \$35	Poor. It is difficult to justify high costs for the uninsured, who make up only 4% of diabetics.	Fairly Good. Capping all insulin costs ensures the same in out-of-pocket costs for all diabetics regardless of insurance status. However, it places the financial burden on taxpayers by nearly doubling government costs.	Fairly poor. Taxpayers are unlikely to support this option because of its high cost. The federal government is unlikely to allocate \$14.24 billion toward helping only 4% of diabetics.
Subsidizing entry costs	Fairly good. This alternative would reduce deaths and high health care spending for uninsured diabetics and provide long-term health benefits. Increasing competition would lower the price of insulin. Subsidizing entry costs would cost \$10 million, which is only 0.00002% of taxpayer revenue. Funding could come from a variety of places including emergency department funding.	Fairly Good. This alternative ensures that every diabetic will eventually have the same out-of-pocket cost for insulin. Increasing competition will lower the price of insulin for everyone. Once out-of-pocket costs reach \$35 per vial, insured diabetics will not have greater access to insulin than uninsured people.	Fair. Stakeholders are likely to accept this option because lowers pressure for pharmaceutical companies, lowers prices for consumers, and improves overall societal health for taxpayers. However, there are costs and administrative burdens associated with changing the current insulin ecosystem. Additionally, pharmaceutical companies may oppose the changes.

the market, increased competition could place sustained downward pressure on insulin prices over time.

See Table 1 for an evaluation of the listed alternatives based on the stated criteria.

Recommendation

Of the alternatives discussed, subsidizing entry costs would provide the best return on investment because it is cost-effective, equitable, and feasible. Maintaining the status-quo would not develop as much equity over time and capping all insulin costs at \$35 is not as feasible. Subsidizing entry costs is the middle ground for all stakeholders. Insulin costs for diabetics will decrease, there will be less of a burden on pharmaceutical companies to manage pricing competition, and taxpayers will see tax revenue being used efficiently now and in the future.

A policy focused on subsidizing entry costs would be implemented through the standard bill passage process in Congress. This requires drafting and introduction, committee review, floor debates, and voting (Keating n.d.). This process takes time and poses a barrier to implementation. Additionally, Congress would need to decide where to allocate the \$30 million from the budget (Keating n.d.). After the policy is approved, existing and potential companies would need to be made aware of the new subsidy. This could also act as a barrier to implementation as companies become aware of the opportunity and consider participating.

Once interested new producers decide to participate, they would apply to receive production approval from the U.S. Food and Drug Administration to ensure they follow appropriate regulations (U.S. Food and Drug Administration n.d.). After they are approved, they would receive \$200,000 to help get them started. New producers would also have the opportunity to apply for \$100,000 for other start-up costs in the first 5 years of insulin production with a cap at \$500,000 spent on each new producer. Acquiring approval from the United States Food and Drug Administration and applying for the subsidy are administrative burdens that new producers would face. Without adequate administrative support, these requirements may discourage market entry. Despite implementation barriers, in the long-term this alternative will create the best insulin

market for the United States by increasing competition, improving price efficiency, and expanding access to insulin.

It is likely that the subsidy policy alternative will receive public and political support. Diabetes has been the eighth leading cause of death in the United States since 2020 (Centers for Disease Control and Prevention 2020). Increasing access to insulin could reduce diabetes-related mortality, making it a policy area with potential bipartisan support. Affordable insulin saves lives and benefits the economy by allowing more people to work and support the economy. Some conservatives may not be on board with increasing welfare for diabetics, or may want to use the \$30 million to support other causes. However, this solution is the best use of money in the long term. If the government invests in increasing competition for insulin producers, there will likely be fewer diabetes related hospital visits, lower costs for taxpayers, and improved health for diabetics. Subsidizing entry costs in the insulin production market will create a more competitive and efficient market, lowering prices, expanding access to life-saving treatment, improving health outcomes, and generating broader economic benefits for society.

References

- American Hospital Association. (2023). Civica Rx gets more backing in effort to reduce drug prices and shortages. <https://www.aha.org/aha-center-health-innovation-market-scan/2023-03-28-civica-rx-gets-more-backing-effort-reduce-drug-prices-and-shortages>
- Baumol, W. J., & Willig, R. D. (1981). Fixed costs, sunk costs, entry barriers, and sustainability of monopoly. *The Quarterly Journal of Economics*, 96(3), 405–431. <https://doi.org/10.2307/1882680>
- Blackstone, E. A., & Joseph, P. F. (2013). The economics of biosimilars. *American Health & Drug Benefits*, 6(8), 469–478. <https://pmc.ncbi.nlm.nih.gov/articles/PMC4031732/>
- Casagrande, S. S., Park, J., Herman, W. H., & Bullard, K. M. (2023). Health insurance and diabetes. National Library of Medicine. <https://www.ncbi.nlm.nih.gov/books/NBK597725/>
- Centers for Disease Control and Prevention. (2020). Diabetes by the numbers. <https://www.cdc.gov/diabetes/about/index.html>
- Centers for Disease Control and Prevention. (2022). About underlying cause of death, 2018–2023, single race. <https://wonder.cdc.gov/ucd-icd10-expanded.html>
- Centers for Disease Control and Prevention. (n.d.). National diabetes statistics report. <https://www.cdc.gov/diabetes/php/data-research/index.html>
- Cha, A. E., & Cohen, R. A. (2020). Reasons for being uninsured among adults aged 18–64 in the United States, 2019. National Center for Health Statistics. <https://www.cdc.gov/nchs/products/databriefs/db382.htm>
- Chesak, J. (2022, March 10). What does insulin cost and what's behind skyrocketing prices? Verywell Health. <https://www.verywell-health.com/insulin-prices-how-much-does-insulin-cost-and-why-5081872>
- Congressional Research Service. (2023). Affordable insulin now act of 2023. <https://www.congress.gov/bill/118th-congress/senate-bill/954>
- Crawford, S. (2023, April 27). The price of insulin: A Q&A with Kasia Lipska. Yale School of Medicine. <https://medicine.yale.edu/news-article/the-price-of-insulin-a-qanda-with-kasia-lipska/>
- DrugPatentWatch. (2025). How nonprofit pharma is reshaping the industry. <https://www.drugpatentwatch.com/blog/rx-for-change-how-nonprofit-pharma-is-reshaping-the-industry/>
- Families USA. (2025). New poll: Crushing health care costs top priority for voters. <https://familiesusa.org/press-releases/new-poll-crushing-health-care-costs-top-priority-for-voters/>
- Fiscal Data. (2024). Federal revenue overview. <https://fiscaldata.treasury.gov/americas-finance-guide/government-revenue/>
- HealthMarkets. (2024, March 14). How much does health insurance cost per month? <https://www.healthmarkets.com/resources/health-insurance/health-insurance-cost-per-mo>
- Johns Hopkins Medicine. (n.d.). Hypoglycemia (low blood sugar). <https://www.hopkinsmedicine.org/health/conditions-and-diseases/diabetes/hypoglycemia-low-blood-sugar>
- Kearsley, A. (2024). HHS standard values for regulatory analysis, 2024. Office of Science and Data Policy. <https://aspe.hhs.gov/sites/default/files/documents/cd2a1348ea0777b1aa918089e4965b8c/standard-ria-values.pdf>
- Keating, B. (n.d.). The legislative process. <https://keating.house.gov/policy-work/legislative-process>

- Lin, Y., Shao, H., Fonseca, V., & Shi, L. (2023). Exacerbation of financial burden of insulin and overall glucose-lowering medications among uninsured population with diabetes. *Journal of Diabetes*, 15(3), 215–223. <https://pmc.ncbi.nlm.nih.gov/articles/PMC10036254/>
- Myerson, R., & Laiteerapong, N. (2016). The Affordable Care Act and diabetes diagnosis and care: Exploring the potential impacts. *Current Diabetes Reports*, 16(4), 27. <https://doi.org/10.1007/s11892-016-0712-z>
- Narendrula, A., Lang, J., & Mossialos, E. (2023). Generic cardiology drug prices: The potential benefits of the Mark Cuban Cost Plus Drug Company model. *Frontiers in Pharmacology*, 14, 1179253. <https://doi.org/10.3389/fphar.2023.1179253>
- Navigator Research. (2025). A majority of Americans oppose cuts to Social Security and Medicare. <https://navigatorresearch.org/a-majority-of-americans-oppose-cuts-to-social-security-and-medicare/>
- Prasad, R. (2019, March 14). The human cost of insulin in America. *BBC News*. <https://www.bbc.com/news/world-us-canada-47491964>
- Rajkumar, S. V. (2020). The high cost of insulin in the United States: An urgent call to action. *Mayo Clinic Proceedings*, 95(1), 22–28. [https://www.mayoclinicproceedings.org/article/S0025-6196\(19\)31008-0/fulltext](https://www.mayoclinicproceedings.org/article/S0025-6196(19)31008-0/fulltext)
- RescueMeds. (2025, December 5). Mark Cuban challenges PBMs with transparency push. <https://rescuemeds.com/mark-cuban-pbm-transparency-reform/>
- Sayed, B. A., Finegold, K. T., Olsen, A., De Lew, N., Sheingold, S., Ashok, K., & Sommers, B. D. (2023). Insulin affordability and the Inflation Reduction Act: Medicare beneficiary savings by state and demographics. Office of the Assistant Secretary for Planning and Evaluation. <https://aspe.hhs.gov/reports/insulin-affordability-ira-data-point>
- Schmalensee, R. (2004). Sunk costs and antitrust barriers to entry. *American Economic Review*, 94(2), 471–475. <https://doi.org/10.1257/0002828041301975>
- Schrier, E., Himmelstein, D. U., Gaffney, A., McCormick, D., & Woolhandler, S. (2024). Taxpayers' share of U.S. prescription drug and insulin costs: A cross-sectional study. *Journal of General Internal Medicine*. <https://pubmed.ncbi.nlm.nih.gov/39446235/>
- Statista. (n.d.). Percentage distribution of household income in the United States in 2023. <https://www.statista.com/statistics/203183/percentage-distribution-of-household-income-in-the-us/>
- U.S. Department of Health and Human Services. (n.d.). Who's eligible for Medicare? <https://www.hhs.gov/answers/medicare-and-medicaid/who-is-eligible-for-medicare/index.html>
- U.S. Food and Drug Administration. (n.d.). The drug development process. <https://www.fda.gov/patients/learn-about-drug-and-device-approvals/drug-development-process>
- United States Senate Committee on Finance. (n.d.). Jurisdiction. <https://www.finance.senate.gov/about/jurisdiction>
- Weilk, E., & Snyderman, R. (2024, January 31). Do you know where your federal taxes go? Bipartisan Policy Center. <https://bipartisanpolicy.org/blog/do-you-know-where-your-tax-dollars-go-tr/>
- World Health Organization. (2024, November 14). Diabetes. <https://www.who.int/news-room/fact-sheets/detail/diabetes>

The European Union's Response to Asymmetric Trade Crises

Distributional Conflict, Institutional Constraints, and the Politics of Unity

Matthew Benjamin¹

Abstract

As the European Commission broadens the focus of European Union trade policy to include market resilience and industrial competitiveness, trade crises impose burdens asymmetrically across member states more frequently and with greater political salience. Given these trends, the Union's ability to act cohesively in the face of these shocks has faltered. While the Commission holds de jure control over EU trade policymaking, recent crises have shown the shortcomings of the Union in maintaining unity among member states on trade matters. This paper argues that the decisive factor shaping EU trade cohesion is the distribution of burdens, both economic and political, across member states. Trade crises which impose broadly symmetrical costs encourage delegation to the Commission from member states; high-impact asymmetrically distributed shocks – particularly in politically sensitive or economically important sectors – expose the limits of existing enforcement and control mechanisms. Through comparing the European Union's response to U.S. steel and aluminum tariffs, Chinese electric vehicle and solar panel subsidies, Ukrainian grain exports, and the Covid-19 pandemic, this paper examines how the distribution of economic burdens shapes the coherence of EU policymaking. It demonstrates that symmetrical impacts often produce unified EU responses, while uneven burden distribution incentivizes unilateral national actions and weakens the Union's position as a global economic actor. The EU's response to the pandemic shows that uneven impacts can still produce unity under exceptional conditions. The paper concludes by identifying structural shortcomings of EU institutions and considers institutional reforms aimed at fostering unity through new policy mechanisms.

Background

Jacques Delors, former President of the European Commission, once described the European Union as an “unidentified political object,” capturing the novelty of its institutions and ambiguity of its political authority. Nowhere is this ambiguity more apparent than in the realm of trade policy. Over recent decades, the Union has emerged as one of the world's most influential trade actors, utilizing its expansive single market, common commercial policy, and regulatory capacity. Yet the EU's effectiveness in trade policy is fundamentally dependent on its ability to act collectively. When the Union is unified, it commands credibility and leverage; when it does not, its credibility and influence are significantly weakened.

Recent trade crises have revealed structural limitations in the EU's ability to maintain cohesion. The Union's unique quasi-federal system, in which member states retain a significant sovereignty while attempting to

coordinate unified trade policies through shared institutions, complicates its ability to respond to trade concerns. While trade policy remains an exclusive EU competence in law, member states have not always complied with Commission preferences nor supported collective responses to economic shocks and trade concerns.

Recent disputes over trade exhibit striking variation in the Union's ability to respond collectively. In some instances, the Union has acted decisively and with unity, with member states delegating authority to the Commission and sustaining a coordinated external posture. In other cases, member states have resisted common measures, made unilateral policy decisions, or openly defied Commission preferences. This variation presents a puzzle: why does the EU respond cohesively to some trade shocks, yet fracture in others?

The importance of trade unity has become increasingly apparent as the Union's trade agenda has broadened. Recent scholarship has

¹ Matthew Benjamin is an undergraduate student at the McCourt School of Public Policy at Georgetown University

emphasized the Union's gradual transition from a primarily liberalizing trade actor towards objectives such as industrial competitiveness, market resilience, and strategic autonomy (Meunier and Nicolaïdis 2019, 103-113). Meunier and Nicolaïdis (2019) characterize this shift as "geopoliticization" and argue that the Commission is at the forefront of defining trade policy's role in pursuing geopolitical aims (Meunier and Nicolaïdis 2019, 105). These ambitions require sustained political cohesion, otherwise risking disunity which undermines Union credibility and leaves all member states vulnerable.

Other scholars discuss the internal constraints of EU trade governance. Dufour (2023) describes the shortcomings of EU institutions, arguing that the decentralization of the EU is a hindrance to Commission objectives and that conflict between member states impedes environmental and social policy objectives (Rubini 2023).

Despite recognition of the shortcomings of EU institutions towards various ends, existing work largely assumes that the Commission can effectively manage internal conflicts through negotiation, exceptions, and limited compensation. The limits of this assumption are seldom assessed in cases of trade shocks that the Commission has managed ineffectively.

This paper argues that the key factor in explaining this variation lies in the distribution of economic and political burdens across member states. When shocks impose relatively symmetrical burdens across member states, or the costs of policy response are shared, member states are generally willing to delegate authority to the Commission and accept collective decision making. When trade crises carry large and highly asymmetric burdens – disproportionately affecting certain member states by impacting industries of national or political importance – EU institutions struggle to manage distributional conflict. In these cases, the Commission lacks sufficient enforcement and compensatory tools to prevent disunity in action, undermining the effectiveness of the common market.

As the European Union seeks to use trade policy as a tool for geopolitical aims, the vulnerabilities presented by disunity become more consequential. Trade crises which impose

uneven costs threaten not only the EU's external credibility, but also the internal stability of the market itself. Understanding why unity breaks down, and how European Union institutions might better manage distributional conflicts, is essential to the health and longevity of the Union's trade regime.

Trade policy is an exclusive competence of the European Union; the Commission is formally empowered to pursue and negotiate trade agreements, enforce trade defense measures, and represent the Union externally. In principle, delegated authority allows the EU to act cohesively as a unified economic actor, enabling coordinated responses to external economic challenges. In practice, however, the durability of Commission measures of this authority depends on the absence of sustained opposition within the Council. While the Commission initiates trade investigations and proposes measures, the adoption and durability of these measures is weak.

Unity in EU trade policy is defined as member states accepting and complying with Commission-led responses, said decisions not being held up by drawn out debate between member states, and member states refraining from unilateral measures. Fragmentation occurs when member states resist, undermine, or defy EU-level decisions through unilateral or ad hoc policy measures on the national level, failing to comply with Commission rulings, or by delaying or blocking EU-level trade responses. Fragmentation does not reflect the absence of de jure Commission authority, but the limits to enforcement and compensatory mechanisms which facilitate collective action.

Under many circumstances, delegating trade authority to the Commission is politically attractive to member states, especially in the face of common threats from external trade actors. Centralized trade policy allows member states to benefit from collective leverage and to avoid bilateral retaliation. When trade shocks impose burdens relatively evenly across member states, unity is maintained through aligned incentives, not coercion.

However, unity breaks down when trade crises generate uneven burdens across the Union. When industrial sectors of national importance or great political power within a few Union members are affected by trade decisions,

governments facing these losses may perceive Commission-led responses as insufficient in addressing national concerns. Furthermore, when shocks benefit certain member states while burdening others – as in the case of an influx of cheap foreign agricultural products hurting agricultural nations while benefiting others – the anticipated stalling of negotiations and insufficient protections granted by the Commission to harmed states also deters delegation. When Commission enforcement mechanisms are weak or compensatory tools are limited, the costs to affected states becomes visible and politically salient, while the benefits of trade unity appear relatively more abstract.

While the Commission possesses *de jure* authority over trade policy, it lacks strong instruments to compel member states into compliance. Trade defense measures, safeguards and temporary holds on trade, and infringement procedures are often slow to implement, induce lengthy litigation, and still often rely on political unity to be imposed. Due to these challenges, the Commission is often forced to fall back on negotiation, temporary holds and exemptions, and informal compromise rather than decisive action and enforcement.

This suggests that EU trade unity is not simply a function of legal competence but also of political economy; symmetric trade shocks bring delegation and unity while large-scale asymmetric burdens expose the fragility of the Union’s centralized trade regime. The following case studies show this dynamic across different trade crises, highlighting conditions which foster unity and shortcomings of current institutions.

Case Study One – U.S. Section 232 Steel and Aluminum Tariffs

The Trump Administration’s U.S. Section 232 steel and aluminum tariffs – implemented between 2018 and 2021 – presented a major external challenge to Europe’s industrial base. The tariff regime imposed 25 percent duties on steel and 10 percent on aluminum as well as additional tariffs on semi-finished and finished industrial products – these tariffs specifically targeted major EU industrial exports (Demertzis and Fredriksson 2018; European Commission 2018). Maria Demertzis and Gustav Fredriksson write in *Intereconomics*, a journal focused on European economic policy, that the 2018 tariffs

targeted roughly €6.4 billion (\$9.61 billion in September 2025 USD) in EU exports and that the United States was “the most important trade partner for the EU” (European Commission 2018).

The scale of American tariffs in 2018 represented a massive shock to EU markets, and this effect was felt across the EU. In 2018, the EU produced 168 million metric tons of steel, representing approximately 9.5 percent of global production – this was concentrated in Germany, France, Italy, and Spain (World Steel Association 2019). Dr. Henrik Adam, the president of the European Steel Association, reflected on the long-term impacts of the Section 232 tariffs, noting, “In 2024 alone, the EU steel industry had to close nine million tonnes of capacity with over 18,000 job cuts announced” (MEPS International 2018). Although these impacts were most visible in Western Europe, they were not confined to those states.

For smaller and less industrial states, these tariffs had cascading effects throughout the entire supply chain for steel. The EU has nine member states producing iron ore, a primary input for steel, including Sweden, Norway, Austria, and Slovakia (AT Minerals 2018). Another primary input, coal, was mined in Poland, Germany, Czechia, the United Kingdom (at the time a member state of the European Union), Greece, Bulgaria, Romania, and Hungary as of 2015 (British Petroleum 2015). U.S. tariffs had effects across the European Union, with supply chain impacts harming nearly every EU country large and small.

The initial reaction of the European Union was a request for consultation at the World Trade Organization (WTO) on June 1, 2018 (Office of the United States Trade Representative 2018). Meanwhile, EU Trade Commissioner, Cecilia Malmstrom, publicly called U.S. measures “pure protectionism” and advocated for their repeal (Rankin 2018). As the Commission and member states negotiated, an important element of the EU’s initial response was officials’ “[insistence that] all EU member states backed Brussels” (Rankin 2018). EU sources claimed that, “All of them (EU member states) feel that at this point in time it is very important to maintain credibility” (Rankin 2018). This demonstrates that delegation to the Commission was viewed as a means of preserving external credibility rather than merely a procedural necessity.

Beyond EU officials bolstering their initial credibility, the European Union actually responded decisively following short negotiations through coordinated legal, diplomatic, and economic countermeasures. The EU exercised Regulation No. 654/2014 which allowed the Commission to amend common commercial policies against third states violating terms of international trade or commitments in the World Trade Organization (WTO). In practice, this allowed the Commission to propose trade enforcement regulation – including targeted tariffs on U.S. products – to the Council for approval from member states via qualified majority voting (QMV).

In June of 2018, the Council passed rebalancing measures in reaction to US tariffs unanimously (European External Action Service 2018). The measures imposed tariffs on U.S. products valued at €2.8 billion (\$4.18 billion in September 2025 USD), including bourbon, textiles, steel and aluminum products, motorcycles, and furniture (EY Global Tax News 2025). Meanwhile, the EU's case in the WTO argued that US tariffs violated GATT and Safeguards Agreement provisions – with the EU as a whole appealing to the WTO rather than individual member states.

The unanimity and solidarity of these decisions is evidence of little national resistance to a collective strategy. The disruption of U.S. tariffs were distributed broadly across EU states, as most were in some way impacted by the supply chains of steel and aluminum products, so all nations had a stake in collective action. Simultaneously, nations supported Annex I (the initial tariff package against US goods) in large part because its provisions were not of particular political or industrial importance to EU countries – most targeted tertiary or finished products, or reciprocal steel and aluminum goods.

Both the initial effects of American protectionism, and the effects of proposed collective action had roughly symmetrical costs across the Union; distributional conflict was minimal as all states were proportionally affected both by the initial challenge and proposed EU response. In this context, delegation to the commission enabled coordinated EU countermeasures without disproportionately harming any single member state's economy.

Case Study Two – Chinese Electric Vehicle Subsidies

Another instance of high-unity is the EU's response to Chinese electric vehicle subsidies. In 2023, the Commission initiated a large-scale anti-subsidy investigation into Chinese electric vehicles (EVs) as concern grew over Beijing's role in EV market distortion. The probe, announced by European Commission President Ursula von der Leyen in September 2023, stated that the Commission had gathered evidence showing that Chinese car manufacturers had benefited from subsidies granted by Beijing, allowing subsidized imports to “rapidly increase their market share in the EU to the detriment of the Union industry” (Brussels Signal 2023). Countervailing measures would depend on the level of subsidization discovered and the extent to which subsidies undercut EU manufacturers. Previous Commission probes into e-bikes and fibre-optic cables identified subsidy margins ranging from 4 percent to 17 percent, providing a reference point for the scale of potential intervention (Nardelli 2024).

The threat Chinese EVs posed to the European market mainly concerned the supply chains of both European EVs and the automotive industry as a whole, as EVs could substitute traditional gasoline-powered cars. Chinese EV prices being “kept artificially low” threatened European car manufacturers concentrated in Germany, Czechia, Spain, Slovakia, and France (European Parliamentary Research Service 2023). Within several EU member states, the automotive sector represents between 1 and 4 percent of national GDP, granting it significant political and economic influence in select countries (European Automotive Industry Association 2021).

But Chinese state-subsidized price undercutting (dumping) consequently affected nearly all constituent parts of the European automotive supply chain. Steel production remains significant in Germany, Belgium, Luxembourg, France, and Sweden, while aluminum is manufactured in Germany, France, Italy, the Netherlands, and Spain (EUROFER 2023). Copper components are produced across the continent, Germany has a prominent plastics and polymer industry, and chassis and subframes are assembled in Central and Eastern Europe (Cognitive Market Research 2025). Internal combustion engine components

are manufactured throughout the Union, while Nordic states have prioritized power cells for EVs and the assembly of higher-end vehicles (Schwierz 2023). Given the complexity and geographic dispersion of automotive production, Chinese EV dumping imposed costs across all EU member states, even if exposure varied by sector and country.

On October 4 2023, the Commission formally launched Case C/2023/160 under EU Regulation 2016/1037, the Union's anti-subsidy regulation. The investigation focused on recently manufactured Chinese EVs, assessing the scale of subsidization and the extent of injury to EU industries. Member states largely deferred to the Commission to resolve this dispute, unifying the European market behind collective action.

Beyond concerns over immediate market distortion, the European Union also saw Chinese dumping as a threat to industrial and economic security (Library of Congress 2023). At the time, few European-manufactured EVs were sold below €30,000, with average prices exceeding €50,000 (Bruegel 2024). Meanwhile, Chinese EVs averaged approximately €32,000 in European markets. With EU nations striving to meet climate and emissions goals, the market share of Chinese-manufactured EVs and low-cost constituent parts – particularly batteries – expanded rapidly (Bruegel 2024).

This market share represented a growing liability for the European Union: as Chinese-made goods accounted for larger percentage of EV sales and constituent parts of European-made EVs, Chinese leverage in a potential trade conflict and the risk of dependency in strategic green industries grew (Bruegel 2024). Chinese state-subsidized goods have benefits, but their drawbacks include “potential public security vulnerabilities, including risks to data and critical infrastructure, and long-term economic dependence, with Europe potentially locked into low-value segments of the EV value chain” (Bruegel 2024). Security also included an elevated risk of corporate espionage and over-dependence on Chinese goods – adding a security dimension which Europe could rally around.

Upon finding evidence of Chinese subsidy, these concerns led the Commission to pre-disclose countervailing tariff rates.

A few sample exporters were named with corresponding prospective tariff rates at 17.4%-38.1% (Euronews 2024). Simultaneously, the Commission opened channels with Chinese authorities to discuss its findings and explore solutions compatible with World Trade Organization rules (European Commission 2024). These legal and diplomatic measures sought to combat the influence of subsidized Chinese EVs in European markets.

The coordinated EU response to American tariffs and Chinese EV subsidies are the types of trade crises the Union is institutionally equipped to manage. Larger trade challenges have been more easily confronted in cases where member states share a common concern. As the objectives of the European Union shift from the promotion of free trade alone toward industrial and geostrategic competitiveness, these cases demonstrate how unity is most easily sustained when member states' interests remain aligned.

Case Study Three – 2022 Ukrainian Grain Crisis

Another set of crises, those with asymmetric burdens on member states and of great-enough scale that negotiation is difficult, represent the greatest threats to European economic unity. These crises make compromise more difficult, and the Union's lack of enforcement mechanisms regarding member states' trade often leads the EU toward disjointed action, leaving it vulnerable to continued disunity.

An example of a crisis demonstrating this vulnerability is the 2022 Ukrainian grain dispute. The outbreak of the Russia-Ukraine War in February 2022 and subsequent breakdown of Black Sea shipping lanes redirected Ukrainian grain exports overland towards Eastern Europe. As maritime exports collapsed, Ukrainian agricultural producers relied heavily on “Solidarity Lanes” established by the EU to move grain by rail, road, and river through neighboring member states. These new imports saturated domestic markets in Poland, Hungary, and other EU member states in the East where logistical bottlenecks prevented onward shipment to third countries.

Farmers in Eastern Europe – a politically sensitive group – effectively pressured their governments to unilaterally impose import and border controls with Ukraine, defying

Commission preferences (Snegovaya and Bergmann 2023). In Poland, where the agricultural sector employs 9 percent of workers, grain prices fell by over 30 percent in some regions between mid-2022 and early 2023, intensifying political backlash (Trading Economics 2026; Sutowski 2023). Ahead of national elections, the crisis became a prominent political issue, and contributed to the party in power, United Right, losing their majority in 2023 (Stanley 2023).

Western European countries, on the other hand, largely experienced the influx of Ukrainian grain as a benefit: consumers appreciated expanded access to grain and storage and transportation remained uncongested. Western member states, as a result, were more willing to prioritize geopolitical solidarity with Kyiv over sectoral protection.

Unity collapsed because trade impacts were unevenly distributed: the Union's East-West divide became far more pronounced as the ability of the Commission to direct trade practices faltered. The geopolitical cohesion of solidarity with Ukraine was insufficient to overcome domestic pressures within member states (European Commission n.d.-a). Beginning in April 2023, Poland, Hungary, Slovakia, and Bulgaria imposed unilateral bans on Ukrainian grain imports, openly contravening EU trade rules and Commission preferences (Snegovaya and Bergmann 2023; Sutowski 2023). The problem was too large and politically sensitive to be solved through simple negotiations, and the Commission lacked strong enough enforcement mechanisms to coordinate member states' actions.

The Commission had insufficient tools available to overcome this crisis and was effectively forced into reactive compromise to placate Eastern Europe. Rather than initiating infringement proceedings or otherwise enforcing compliance, the Commission negotiated temporary measures to protect farms and organized small financial compensation packages. Eventually, grain corridors were negotiated to export Ukrainian grain from Polish and EU ports to third nations, but the Commission failed to organize decisive action in face of these challenges. The incident revealed the limits of the EU's trade authority when asymmetric shocks intersect with electoral politics and highly mobilized domestic

constituencies; the EU faltered when confronted with national opposition to decisions made by the Commission.

Case Study Four – Chinese Solar Panel Subsidies and Dumping Practices

Another crisis which threatened European unity was the response to Chinese solar panel subsidies and dumping practices in the early to mid-2010s. A surge of cheap solar panel imports in the early 2010s posed an existential threat to European solar manufacturing firms, especially in Southern Europe, launching the largest trade defence case in the Union's history. Between 2010 and 2013, the EU's global market share of solar panel manufacturers collapsed as Chinese imports – subsidized by Beijing – undercut European producers.

This trade challenge represents high yet unevenly distributed industrial exposure, and the kind of crisis which uniquely challenges the EU's quasi-federal model. Although the Commission moved to impose provisional anti-dumping duties, support across member states was divided. In 2012, the Commission initiated a formal anti-dumping investigation which concluded that Chinese solar panels were being sold in the EU at prices far below fair market value (European Commission 2013). The Commission's initial anti-dumping duties on Chinese solar technologies was set far below the assessed dumping margin, yet many member states opposed the tariffs. Germany, the United Kingdom, the Netherlands, and other countries heavily reliant on solar panel imports opposed strong punitive measures (CNBC 2013).

The geopolitical goals of the Commission, in this case a commitment to the green transition and the goal of protecting and promoting European industry, came to a head. The harm to European solar manufacturers was localized in Southern Europe – particularly Italy and Spain – and the negotiated price undertaking and temporary duties failed to restore a competitive European solar panel industry. In the following years, many European solar manufacturers went bankrupt or relocated; as these manufacturers left the market, downstream installation sectors continued to expand (Green European Journal 2020).

The Commission's strongest trade-defense instruments, anti-dumping measures and tariffs, are highly constrained by large-scale distributional conflict. The Commission lacks the tools to properly protect or compensate localized European industries and the Commission's trade policy reflects an orientation and reliance on compromise rather than a common commitment to robust industrial protections. In the absence of better compensatory mechanisms, the costs of Commission decisions were concentrated while benefits were more broadly distributed. The inadequacy of Commission action to protect or otherwise compensate locally-concentrated sectors effectively doomed them financially.

This case illustrates the insufficient strength of EU-level trade policymaking. The fragmentation weakened not only European solar manufacturers, but also signaled to foreign industries and competitors that the EU may not stand firm when faced with similar future challenges. There is a structural vulnerability for trade exploitation when member states' domestic economic interests diverge. Although the fallout from Chinese solar panel dumping contributed to subsequent expansions of EU trade defense mechanisms – allowing the Union to more-aptly respond to Chinese EV subsidies – the Union still lacks tools to effectively combat dumping in sectors with asymmetric burdens and large localized losses.

Case Study Five – COVID-19 Pandemic

The European Union has previously been able to implement new Commission tools in times of crisis, most notably during the COVID-19 pandemic. This response is the clearest contemporary example of the Union's ability to expand Commission authority, and mobilize joint financial resources which were previously politically unavailable. Rather than relying on existing competences, the pandemic prompted institutional innovation which expanded supranational authority. The pandemic, in its early stages, asymmetrically affected the Union, with Italy and parts of Southern Germany being hit early on, yet the Union's eventual response demonstrated its capacity to centralize authority to mitigate highly uneven shocks (Siddi 2020). This offers

a relevant precedent for how the EU might address future asymmetric trade crises.

Initial state responses of closed borders and export controls on medical supplies threatened the health of the Union; but the depth of the crises forced member states to delegate powers to the Commission to regulate supply chains, coordinate procurement of medical supplies, and ensure cross-border flows of essential goods and services (Baccaro and Blyth 2024; European Commission 2020). The Recovery and Resilience Facility (RRF) allowed the Commission to borrow collectively on capital markets, distribute grants and loans to member states, tie funding to EU-level policy priorities, and exercise limited oversight in national spending plans (European Commission 2025b). The Commission gained temporary, but meaningful, fiscal and coordinating authority which went beyond its traditional role in economic governance.

This shift of the Commission, from a weakened executive arm of the EU to having the power to coordinate funding and crisis response, showed how powers can be used to effectively mitigate crises. More than enforcement powers, compensatory and redistributive mechanisms helped offset costs and aid states disproportionately affected by the pandemic. With this precedent, the EU may benefit from gradually shifting some of these powers towards dealing with large-scale trade challenges and allowing the Commission to more effectively coalesce members to respond to trade concerns.

The EU was able to create effective new fiscal instruments which had compensatory and redistributive mechanisms among member states on an unprecedented scale. The deployment of these tools proves their feasibility within the EU; although recent challenges have exposed shortcomings in the current Commission's toolkit, the pandemic response suggests that institutional constraints are not purely legal, but political. Creating analogous trade policy mechanisms to those used in the EU's pandemic response represents an extension of existing crisis-management rather than an unprecedented expansion.

Discussion

The European Union's economic diversity and decentralized production structure ensure that trade shocks seldom fall evenly across member states. Supply chains, industries, and agricultural sectors remain regionally concentrated. Goods with complex, Europe-wide supply chains – such as automobiles and finished metal products – transcend this localization, but primary and tertiary goods imported to Europe affect certain countries and regions far more than others.

When shocks impact major industries or politically sensitive groups, domestic political pressures quickly override incentives for EU-level coordination. Governments facing concentrated losses among farmers, industrial workers, or strategically important firms have costs of compliance with Commission-led trade policy which become immediate and visible; the benefits of market unity appear distant or abstract. Under this incentive structure, national authorities often pursue unilateral and ad hoc measures even when such measures undermine the integrity of the common market.

The Commission's limited coercive capacity intensifies this dynamic, creating structural vulnerabilities in EU trade policymaking as the Union embraces geoeconomic and industrial goals. The frequency and salience of asymmetric shocks is expected to increase as foreign competitors exploit these institutional gaps and member states specialize in different sectors. Without tools capable of managing distributional conflict at scale, the EU risks repeated periods of fragmentation that weaken both internal cohesion and external credibility.

The European Union now stands at a crossroads. Vulnerability arising from the structure of current trade institutions expose the Union to repeated fragmentation. The Union must choose to pursue deeper integration – granting the Commission greater enforcement powers on trade matters in the short and long terms – or resign to a system of tiered integration or selective cooperation.

The punitive actions available to Brussels often amount to a cost of doing business for member states rather than a credible deterrent. In practice, the Commission's tools are politically bottlenecked, constrained by strict criteria for deployment, prone to stalling

negotiations among member states, and are frequently imposed with time lags which reduce their effectiveness.

Trade defence instruments (TDIs), including anti-subsidy safeguards, are subject to a "Union interest test," which requires that a proposed tariff or countermeasure serves to benefit the overall EU rather than specific industries or regions. This requirement creates a structural constraint: imports that impose significant harm on particular member states or sectors may still be permitted – or subject to diluted remedies – if they generate some benefits to the EU. When this pattern persists, especially to the detriment of certain member states or industries over others, the incentives for compliance erode and the legitimacy of centralized trade governance weakens.

Safeguard measures, including quotas and temporary restrictions, face similar limitations. Levied in cases of sharp, unforeseen, and serious surges in imports which threaten EU industry, the deployment of safeguards, again, necessitates the "interest of the Union." The high threshold for activation, combined with lengthy procedural requirements, diminishes the Commission's ability to appropriately respond to trade developments. Therefore, historically, safeguards have been deployed "sparingly" (European Parliamentary Research Service 2025; European Commission n.d.-c).

More recently introduced instruments – approved since 2019 – remain legally sensitive, under used, and unproven. These instruments represent a positive development for European trade unity, adding a second layer of tools beyond classical trade defence for the Commission, yet their de jure provisions remain seldom enacted. The ability of the Commission to investigate companies participating in distortive subsidies, and block their mergers or public procurement in the EU was passed in 2023 (Parlementaire Monitor 2020). While these tools add a second layer of authority beyond traditional TDIs, they rely on qualified majority voting and can be delayed or diluted for months should any strong opposition present itself. Lowering the thresholds for using these tools, or establishing conditions for their automatic deployment, may strengthen the Union more effectively than creating new trade instruments with similarly limited enforceability.

Foreign direct investment (FDI) screening regulations, meant to secure strategic industries from foreign interference or distortive foreign investment, relies on national implementation – causing fragmentation among member states lacking central authority (European Commission n.d.-d; Council of the European Union n.d.). These tools are an appropriate move towards securing the geopolitical aims of the Commission, yet remain completely ineffective when member states' interests diverge. While the Commission possesses an expanding legal mandate, it lacks the capacity to act decisively or keep up with an evolving trade landscape.

Moving forward, should the EU and its members choose further integration to resolve current structural shortcomings, both short and long term policy changes must be enacted.

In the short term, the EU can create new, rules-based compensation mechanisms available to member states disproportionately harmed by trade shocks. This compensatory tool may be formally embedded within the European Union's budget or integrated into existing calculations for member-state benefits and contributions. Creating these systems would better allow the Commission to buy member-state compliance during asymmetric crises and mitigate the impacts of trade burdens; automatic adjustments for trade burdens also prevents EU responses from being held up in case-by-case litigation or prolonged intergovernmental bargaining.

The passage of EU automated trade adjustment and compensation tools may still be held up by countries expected to become net contributors, and any state anticipated to benefit from adjustments may attempt to game the system; however, programs with similar objectives have proven feasible institutional settings.

Trade Adjustment Assistance (TAA) in the United States was established in 1962 and expanded via the Trade Act of 1974 and subsequent reauthorizations. TAA is designed to help workers displaced by import competition; it provides job retraining and job-search assistance, temporary income support, and relocation subsidies (U.S. Department of Labor n.d.). There is also a firm-level variant (TAAF), which may be more applicable to EU member states with industries of national

importance, that offers funds to firms to help them restructure and regain competitiveness. Empirical evaluations of TAAF indicate that participating firms have, in a significant number of cases, restored sales, productivity, and employment following adjustment assistance (U.S. Economic Development Administration 2018; U.S. Economic Development Administration 2017). These kinds of targeted compensatory investments, applied to firms or adapted for the EU member-state level, demonstrate that compensation and adjustment schemes are feasible and can be applied to trade shocks.

The European Globalization Adjustment Fund (EGF) was created in 2006 with the goal of supporting workers laid off due to trade liberalization and global economic changes. The program provides similar services to TAA, adding micro-credits and entrepreneurial support, but is tailored narrowly to workers in firms experiencing mass layoffs (European Commission n.d.-b). This tool remains limited in scale and has never been adjusted to become a structural tool. The EGF exemplifies EU experimentation with compensatory tools tied to trade shocks, but its restricted scope means that it does not approach the robust automatic adjustment and compensation tools necessary for larger geoeconomic challenges.

Additionally, in the short term, formalizing and enhancing emergency delegated powers of the Commission – as seen in pandemic-era joint procurement and coordination – would strengthen the Union's ability to act quickly and cohesively. While these powers are likely to remain underutilized, their formalization and stated intent to apply to large-scale trade crises increase the Commission's credibility in resolving trade concerns.

In the long term, it must be a sustained objective of the Union – should deeper integration and clearer resolution of asymmetric trade concerns be desired among member states – to develop an integrated economic security governance structure capable of effectively binding member states to collective action. Strengthening the enforcement capacities available to the Commission, possibly through Treaty change, is necessary to ensure that unilateral national trade responses become legally, economically, or politically too costly to pursue. This can be accomplished by expanding

the legal jurisdiction and powers of the European Union, pairing trade customs enforcement with broader external border and future defense cooperation, and expanding Brussels' ability to apply punitive economic measures – such as redistributing aid or levying financial penalties against member states in defiance of Commission rulings. All enforcement capacity must be made with the intent of actually binding states' behavior rather than adding a cost of doing business; absent credible deterrence, disunity is likely to continue.

These integrations and enforcement mechanisms have not already been expanded or formalized due to persistent member-state hesitancy towards greater centralization in Brussels. Trade policy, while an exclusive EU capability, remains entangled with national industries and politics. Many governments hesitate to empower the Commission, fearing that stronger coercive tools may be deployed to the detriment of nationally important industries in future crises. Net-contributor states have resisted institutionalizing compensatory mechanisms which could impose additional fiscal obligations, while other states fear setting additional precedents for supranational interventions beyond temporary crisis responses. As a result, EU integration has occurred incrementally and often as a reaction to world events rather than being codified *ex ante*.

A wider goal of EU policymaking – deepening fiscal and industrial integration and cooperation – is a complementary trend for mitigating severe trade concerns. With greater transnational integration, the political salience of asymmetric costs may be reduced by extending supply chains across member states and diffusing exposure across national economies. European successes in combating Chinese EV dumping and American steel and aluminum tariffs are attributable, in part, to deeply integrated, pan-European supply chains for related goods. This is a long-term solution which would gradually reduce asymmetries in the burdens of trade – especially for secondary and tertiary products.

Deeper industrial integration will not solve all market challenges, especially those concerning primary goods such as grain, foodstuffs, and raw minerals, and therefore must be developed alongside additional tools

of the Commission. Lowering the thresholds for deploying existing tools may abate these concerns, but doing so could intensify political deadlock and provoke resistance from the Union's larger economies. The EU will not effectively combat asymmetric trade shocks without institutional mechanisms capable of enforcing collective actions when incentives to defect emerge.

Conclusion

The EU's future as a geoeconomic actor depends on its ability to manage trade crises that unevenly affect its member states. As the Union pursues geopolitical and industrial objectives beyond classical free trade in its trade policymaking, high-asymmetry shocks will become more frequent and politically salient. The Union's capacity to facilitate collective trade action in response to these kinds of shocks has remained inadequate, leaving a structural vulnerability in the European Union's trade governance.

This vulnerability threatens the cohesion of the Union itself. When the Commission is perceived as unable to adequately address nationally-concentrated concerns, member states face strong incentives to act unilaterally, even at the expense of the common market and EU credibility. Such fragmentation undermines EU-level governance and creates opportunities for external actors to exploit division between member states and weaken the Union's collective power.

Absent effective countermeasures to manage asymmetric burdens, trade crises risk eroding the European Union as a unified economic actor. However, this outcome is not inevitable. By recognizing asymmetrical trade shocks as a distinct and recurring challenge facing the union, and by adopting reforms tailored to combat them despite political division, the European Union can bolster the single market, preserve internal solidarity, and strengthen its role as a global economic power.

References

- AT Minerals. “Metal Ore Mining in Europe.” 2018. https://www.at-minerals.com/en/artikel/at_Metal_ore_mining_in_Europe-3257608.html
- Baccaro, Luca, and Mark Blyth. “Industrial Policy and European Integration.” *Journal of European Public Policy* 31, no. 1 (2024). <https://www.tandfonline.com/doi/full/10.1080/13501763.2023.2294143>.
- Bloomberg. Nardelli, Alberto. “EU Moves toward Hitting China with Tariffs on Electric Vehicles.” March 6, 2024.
- British Petroleum. *Statistical Review of World Energy 2015*.
- Bruegel. “A Smart European Strategy for Electric Vehicle Investment vis-à-vis China.” 2024. <https://www.bruegel.org/policy-brief/smart-european-strategy-electric-vehicle-investment-china>.
- Brussels Signal. “EU Formally Opens Anti-Subsidies Probe into EVs Made in China.” October 2023. <https://brusselssignal.eu/2023/10/eu-formally-opens-anti-subsidies-probe-into-evs-made-in-china/>.
- Center for Strategic and International Studies. Snegovaya, Maria, and Max Bergmann. “Fracturing Solidarity: The Grain Trade Dispute between Ukraine and the European Union.” 2023. <https://www.csis.org/analysis/fracturing-solidarity-grain-trade-dispute-between-ukraine-and-european-union>.
- Cognitive Market Research. “Vehicle Subframe Market Report.” <https://www.cognitivemarketresearch.com/vehicle-subframe-market-report>.
- Council of the European Union. “EU Trade Legislation.” <https://www.consilium.europa.eu/en/policies/eu-trade-legislation/>.
- Demertzis, Maria, and Gustav Fredriksson. “The EU Response to US Trade Tariffs.” *Intereconomics* 53, no. 5 (2018): 253–258. <https://www.intereconomics.eu/contents/year/2018/number/5/article/the-eu-response-to-us-trade-tariffs.html>.
- Electrive. Schwierz, Peter. “Volvo and Northvolt Start Building Battery Cell Factory in Gothenburg.” September 28, 2023. <https://www.electrive.com/2023/09/28/volvo-and-northvolt-start-building-battery-cell-factory-in-gothenburg/>.
- EUROFER (European Steel Association). *European Steel in Figures 2023*. Brussels: EUROFER, 2023. https://www.eurofer.eu/assets/publications/brochures-booklets-and-factsheets/european-steel-in-figures-2023/FINAL_EUROFER_Steel-in-Figures_2023.pdf.
- Euronews. “What Brussels Found When It Probed Chinese Subsidies and Electric Cars.” June 13, 2024. <https://www.euronews.com/my-europe/2024/06/13/what-brussels-found-when-it-probed-chinese-subsidies-and-electric-cars>.
- European Automotive Industry Association. “Position and Perspectives of the European Automotive Industry.” 2021. <https://www.researchgate.net/publication/355144184>.
- European Commission, “Anti-Dumping Duties on Solar Panels from China.” Press Release IP/13/501, June 4, 2013. https://ec.europa.eu/commission/presscorner/detail/en/ip_13_501.
- European Commission, “Commission Pre-Discloses Countervailing Duties on Battery Electric Vehicles from China.” Press Release IP/24/3231, 2024. https://ec.europa.eu/commission/presscorner/api/files/document/print/en/ip_24_3231/IP_24_3231_EN.pdf.
- European Commission, “Commission Unveils €540 Billion EU Budget Support Measures for COVID-19 Recovery and Resilience.” Press Release IP/20/523, May 27, 2020. https://ec.europa.eu/commission/presscorner/detail/en/ip_20_523.
- European Commission, “EU Response to US Tariffs on Steel and Aluminum: Questions and Answers.” 2018. https://ec.europa.eu/commission/presscorner/api/files/document/print/en/qanda_25_750/QANDA_25_750_EN.pdf.

- European Commission, "EU Solidarity with Ukraine." n.d.-a. https://commission.europa.eu/topics/eu-solidarity-ukraine_en.
- European Commission, "European Globalisation Adjustment Fund for Displaced Workers (EGF)." n.d.-b https://employment-social-affairs.ec.europa.eu/policies-and-activities/funding/european-globalisation-adjustment-fund-displaced-workers-egf_en.
- European Commission, Joint Research Centre. "Aluminium." Raw Materials Information System. 2025a. <https://rmis.jrc.ec.europa.eu/rmp/Aluminium>.
- European Commission, "Recovery and Resilience Facility." October 8, 2025b. https://commission.europa.eu/business-economy-euro/economic-recovery/recovery-and-resilience-facility_en.
- European Commission, "Safeguard Measures." n.d.-c. https://policy.trade.ec.europa.eu/enforcement-and-protection/trade-defence/safeguards_en
- European Commission "Trade Enforcement and Protection." n.d.-d. https://policy.trade.ec.europa.eu/enforcement-and-protection_en.
- European External Action Service. "EU-US Trade Relations." June, 2018. https://www.eeas.europa.eu/node/46980_en.
- European Parliament, "EU Trade Defence Instruments: Recent Developments." 2025. [https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI\(2025\)775919](https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2025)775919).
- European Parliamentary Research Service. "EU Anti-Subsidy Investigation into Chinese Electric Vehicles." 2023. [https://www.europarl.europa.eu/thinktank/en/document/EPRS_ATA\(2023\)754553](https://www.europarl.europa.eu/thinktank/en/document/EPRS_ATA(2023)754553).
- EY Global Tax News. "EU Responds to US Tariffs on Steel and Aluminum." 2025. <https://globaltaxnews.ey.com/news/2025-0689-eu-responds-to-us-tariffs-on-steel-and-aluminum>.
- Green European Journal. "Making Europe's Solar Future." 2020. <https://www.greeneuropeanjournal.eu/making-europes-solar-future/>.
- Guardian, The. Rankin, Jennifer. "EU Starts Retaliation against Donald Trump's Steel and Aluminium Tariffs." June 1, 2018. <https://www.theguardian.com/business/2018/jun/01/eu-starts-retaliation-against-donald-trumps-steel-and-aluminium-tariffs>.
- Intereconomics. Demertzis, Maria, and Gustav Fredriksson. "The EU Response to US Trade Tariffs." 2018.
- IPG Journal. Sutowski, Michał. "Seeds of Strife: How the War in Ukraine Reshaped Europe's Agricultural Politics." 2023. <https://www.ips-journal.eu/topics/economy-and-ecology/seeds-of-strife-7361/>.
- Library of Congress. "European Union: Commission Launches Anti-Subsidy Investigation into Imports of Battery Electric Vehicles from China." Global Legal Monitor, November 1, 2023. <https://www.loc.gov/item/global-legal-monitor/2023-11-01/european-union-commission-launches-anti-subsidy-investigation-into-imports-of-battery-electric-vehicles-from-china>.
- LSE Europb Blog. Stanley, Ben. "What We Learned from the 2023 Polish Election." October 24, 2023. <https://blogs.lse.ac.uk/europpblog/2023/10/24/what-we-learned-from-the-2023-polish-election/>.
- MEPS International. "EU Steelmakers Eye 2018 to Determine US Tariffs Impact." <https://mepsinternational.com/gb/en/news/eu-steelmakers-eye-2018-to-determine-us-tariffs-impact>.
- Meunier, Sophie, and Kalypso Nicolaïdis. "The Geopoliticization of European Trade and Investment Policy." *Journal of Common Market Studies* 57, no. 1 (2019): 103–113. <https://kalypsonicolaïdis.com/wp-content/uploads/2020/11/jcms.12932.pdf>.

Office of the United States Trade Representative. "Certain Measures on Steel and Aluminum Products." <https://ustr.gov/issue-areas/enforcement/dispute-settlement-proceedings/wto-dispute-settlement-pending-wto-disputes/certain-measures-steel-and-aluminum-products-1>.

Parlementaire Monitor. "EU Trade Policy and Institutional Authority." <https://www.parlementairemonitor.nl/9353000/1/j9vvij5epm-j1ey0/vk0gkfd9vbfw>.

Rubini, Luca. "Trade Policy, Economic Security, and EU Unity." *Politics and Governance* 11, no. 2 (2023). <https://www.cogitatio-press.com/politicsandgovernance/article/view/6174>.

Siddi, Marco. *Italy and the COVID-19 Emergency*. Helsinki: Finnish Institute of International Affairs, 2020. <https://fia.fi/en/publication/italy-and-the-covid-19-emergency>.

"Trade Adjustment Assistance for Firms Success Stories." 2017. <https://www.eda.gov/archives/2021/news/blogs/2017/03/01/highlight.htm>.

Trading Economics. "Poland: Employment in Agriculture (% of Total Employment)." <https://tradingeconomics.com/poland/employment-in-agriculture-percent-of-total-employment-wb-data.html>.

U.S. Department of Labor. "Trade Adjustment Assistance." <https://www.dol.gov/general/topic/training/tradeact>.

U.S. Economic Development Administration. "How the TAAF Program Is Helping U.S. Manufacturing Firms." <https://www.eda.gov/impact/success-stories/manufacturing/examples-how-taaf-program-helping-us-manufacturing-firms>.

World Steel Association. "Global Crude Steel Output Increases by 4.6% in 2018." 2019.

THIS PAGE LEFT INTENTIONALLY BLANK.

The Future of Work: The Case for a Robot Tax

Nikola D. Misetić¹

Introduction

The rapid growth and sophistication of artificial intelligence over the last three years has already started a massive reshaping of global markets. What makes AI unique from prior automation transformations is the type of work it is able to replace. While previous advancements, like the assembly line, primarily transformed manual labor, AI is becoming increasingly good at substituting cognitive tasks. This technological transition encompasses a wide array of logistical, legal, and creative functions across many diverse professional sectors. The near-universal need for cognitive function in the workplace means that modern AI systems elevate the risk of widespread job displacement across both blue and white collar professions.

While many advocates claim that AI will bring significant productivity gains to the U.S. economy, the government currently has no strategy to address the consequences for when automation replaces human labor. This has led to about 52% of American adults feeling concern rather than excitement about the use of AI (West et al. 2025). The federal government's position on this issue leans in favor of market-led adaptation and self-correction, but provides limited mechanisms for internalizing labor-displacement costs at the firm level (Kratisios et al. 2025). This also increases the risks of severe labor adjustment lags, which could lead to serious long-term unemployment, economic decline, and the erosion of public trust (Acemoglu & Restrepo 2018). Current estimates suggest that AI can replace up to 25% of current work tasks in the U.S. and Europe (Hatzius et al. 2023). With such a large proportion of the workforce at risk of being laid off, it is imperative that the government has a strategy to avoid, or at least mitigate, this outcome.

This paper argues that the U.S. government should adopt a target Automation Adjustment Contribution, which is more colloquially known as a "robot tax." This tax would apply to large firms who simultaneously increase task

automation and reduce net employment beyond a measurable displacement threshold. The revenue generated by the robot tax would go towards job retraining programs for individuals affected by AI automation, which would be managed by the Department of Labor (DOL) and other agencies. This paper also recommends a five year pilot program for the robot tax in order to measure its effectiveness before implementing it at a national scale. Promising pilot program results will demonstrate that a robot tax can offset job displacement while still encouraging growth, as well as transform AI automation into the foundation for economic progress.

Economic Justification

AI driven replacement of labor is a textbook externality. Firms reap the benefits of automation through reduced costs and increased output, yet bear none of the responsibilities for negative consequences such as long-term unemployment and economic contraction. Free market incentives alone are unable to encourage firms to reskill displaced workers, which is why state programs are necessary to maintain labor liquidity and prevent demand-side economic weakness (Acemoglu & Restrepo 2018). Furthermore, efforts to internalize costs have shown success in the past. Where carbon pricing internalized environmental costs without reducing innovation, a robot tax would internalize labor-displacement costs without reducing automation.

A robot tax should only take effect when a firm uses AI to simultaneously substitute labor and reduce employment, not just when it is used at all. This avoids punishing firms that use productivity tools, like Microsoft Excel, that assist rather than replace workers. Unlike AI systems that utilize machine learning to simulate human reasoning, Excel is a deterministic software application that serves strictly as a static utility tool for manual data organization. This distinction is what separates the robot tax from a responsible substitution policy proposal

1 Nikola D. Misetić is an MPP student at the McCourt School of Public Policy at Georgetown University

and a blanket tax on technology. Measured labor transitions preserve tax bases and fuel long-term productivity by using automation dividends to fund human capital (Gallant 2025). Using this framework, the robot tax ensures that innovation remains economically, efficient, politically legitimate, and socially sustainable.

Methodology

To ensure that our robot tax is grounded in empirical research rather than speculative theory, the policy will utilize a dual-metric framework to identify when it should be triggered. This framework moves beyond a “blanket technology tax” by focusing specifically on the correlation between capital investment in AI and net labor reduction. These metrics are critical for policy validation and for establishing a technical basis of taxation that is based on the statistical risks of AI replacement (Hatzius et al. 2023). Ultimately, a data-driven approach will improve the ability of the robot tax to properly internalize the costs of AI automation.

The primary tool for measuring automation intensity will be the “task substitution rate” as defined by the International Labor Organization (ILO). According to the ILO, task substitution rate effects will be most prevalent in clerical support professions first and disproportionately impact women, estimating roughly 2.3 to 5.1% of employment being at risk of automation (Berg 2025). When designing the robot tax threshold, the Department of Commerce (DOC) should utilize the ILO’s established metrics and research to identify the share of human tasks now performable by AI systems. Using this threshold, a firm will only be subject to the robot tax if its adoption of AI targets “substituting” tasks rather than “assisting” tasks, preventing the taxation of productivity tools.

The second component of the robot tax needs to address the “correlation vs. causation” debate and prove that automation is the direct driver of displacement at a given firm. To do that, the tax needs to employ different triggers that respond to statistical measurements. The first would be a “net employment threshold,” which triggers the tax when a large firm (100+ employees) experiences a 10% contraction in “high-risk job categories” despite increasing AI-task density (Metzenbaum 1988; Cook 2024). The DOL will then examine further and

calculate the firm’s “Automation Displacement Ratio” using the following formula:

$$ADR = \frac{\Delta \text{ Employees in AI Exposed Roles}}{\Delta \text{ Total AI Task Substitution Rate}}$$

Finally, implementing the robot tax must include a constant data-collection mechanism in order to satisfy the need for evidence-based governance. The DOC, for instance, would establish annual reporting guidelines that track correlations between AI expenditures and payroll savings to determine if any layoffs were a strategic result of technological substitution rather than general market volatility (Cazzaniga et al. 2024). Additionally, the Government Accountability Office (GAO) will be tasked with measuring the effectiveness of the employee retraining programs funded by the tax. Their assessment would be based on three different quantitative variables: reemployment rates, retraining program participation, and employer compliance levels.

Policy Benefits

The primary benefit the robot tax will have on the economy is stabilizing the labor market. As discussed earlier, public intervention is necessary to address automation externalities. With the money generated from a robot tax, we can establish and fund job reskilling programs to aid in labor transitions. This way, companies can now externalize reskilling costs rather than internalize them. Without this intervention, workers will struggle to access similar programs, reinforcing inequality and reducing their competitiveness in a labor market increasingly defined by AI literacy (Lin & Wu 2025).

In addition to protecting the labor force, a robot tax will also improve our overall economic resilience. In this case, economic resilience refers to absorbing labor shocks while preventing damage to local communities. When large employers automate and reduce headcount, small businesses suffer and consumer tax brackets shrink. A robot tax helps stabilize local labor markets during periods of transition by reducing regional decline and labor force detachment (Kamal-Chaoui 2025). In macroeconomic terms, it prevents automation induced productivity from being offset by consumption slowdowns.

With a strong labor force and a resilient economy, a robot tax is a great method for

improving global competitiveness. Countries that are adopting joint innovation and resilience programs are set to outperform traditional laissez-faire strategies. Singapore and Germany, for example, have demonstrated that modern competitiveness requires national human investment that matches automation scaling (Zadrezwska 2025). Furthermore, evidence from the U.K. and China suggests that firms that integrate AI with “robust support” are more resilient during economic slowdowns and better at managing crises like the COVID-19 pandemic (Drydakakis 2025). Without reskilling funding, the U.S. risks creating a skills gap that undermines both productivity and its position as the global AI leader.

Finally, a robot tax would be the perfect way to address worldwide concerns over responsible AI usage. The idea that innovation should pay for its own externalities is one with precedence in the U.S. through the carbon pricing program. Mechanisms like the Cap-and-Trade program, for instance, generate funds that go toward climate-related programs while also providing financial incentives for firms to innovate (Randolph 2025). Regarding the AI sector, South Korea currently leads in internalizing negative externalities after becoming the only country to have any kind of robot tax (Jacinto 2023). By joining South Korea, the U.S. will secure its position as a predictable and stable environment for AI innovation and investment, thereby establishing itself on the global stage as an AI pioneer.

Addressing the Counterarguments

“Markets will Self-Correct”

Proponents of the free market believe that most if not all government intervention in industry will only hurt the natural equilibrium. These people say that AI will create job transformation rather than displacement—producing new sectors over time while eliminating old ones (Frazier 2026). Using this perspective, short-term disruptions to the labor market are preferable to long-term stagnation caused by government intervention. Proponents also note that intervention has had mixed success in the past and will only burden the tax payer unnecessarily.

While it is true that markets will and often course correct, the issue at hand is whether

or not that will occur before severe labor displacement takes place. The unique issue with the AI wave comes from its unique ability to replace cognitive skills. The market cannot guarantee equivalent job transfer as a result of the skill mismatch that will exist within an AI transformed job market. These hypothetical new jobs would require a high proficiency in AI systems, while the substituted workers will lack the means to retrain themselves (Benhamou 2023). A robot tax would therefore accelerate the market’s self-correction processes by funding reskilling programs that can better prepare the workforce.

“Defining Automation is Complex”

Critics of regulation in general often say that they place administrative burdens on industry without considering the economic consequences. In the case of AI, critics claim that any proposed regulation on “automation replacement” is inherently vague (Ebbin 2026). Automation in the workplace can vary significantly—physical robots can replace factory workers, while AI copilots can take on paralegal labor. This variation illustrates how regulations need to understand the difference between AI assisting workers and substituting them. If automation is measured too broadly, efficiency tools that greatly enhance productivity would be subject to the tax.

While this challenge is certainly real, it is completely solvable. Both national and international labor bodies have already begun testing substitution metrics in the context of new AI tools. The International Labor Organization, for example, is measuring the share of human tasks now performable by machines as the task substitution rate (Samman 2022), as well as net changes in employment due to technology adoption (ILO 2025). The DOL and other agencies also collect information on job displacement that can be adapted to define automation thresholds. We can use these data and definitions to create baseline thresholds for future taxes and regulations.

“Robot Taxes Kill Innovation”

One of the most popular arguments against taxes overall is that they stifle innovation by reducing competitiveness. Large AI firms may choose to move their operations overseas without the tax, reducing GDP (Mann 2019).

Additionally, small AI firms may not be able to compete, or even form, if they cannot afford to pay the tax (Heer 2026). Critics conclude that the tax's potential to stifle innovation will lower our economic output and reduce national power.

The issue with this argument is that it assumes that the tax will freeze automation completely and that firms will immediately pick up and leave. The robot tax is not a development tax, but an adjustment tax that only triggers when AI substitution causes a disproportionate workforce reduction. Additionally the tax includes “smart” thresholds for firm size, payroll savings, and net displacement in order to protect startups as they grow. Furthermore, the political instability and distrust caused by a potential labor market disruption would do more harm to the AI industry and innovation than a tax on the highest earning firms (West et al. 2025).

Implementation Strategy

Before attempting to broadly implement a robot tax, it is important to conduct a pilot test to see what the outcomes of the policy will be and find areas for improvement. Our pilot program would apply a robot tax on industries where we would see the greatest immediate impact, such as manufacturing, warehousing, and transportation. The timeline for the pilot project would be structured as follows:

Year 1: Administrative Framework and Program Funding

The robot tax would be administered through an interagency framework that takes advantage of their existing federal capacities rather than creating a new bureaucracy. The Internal Revenue Service (IRS) would be tasked with collecting the revenue associated with qualifying automation levels per firm. Next, the DOL would facilitate the transfer of these funds toward an “AI Worker Transition Program,” partnering with state agencies to ensure said funds are deployed properly. Finally, the DOC would create an impact assessment to establish reporting requirements and provide a technical basis for when the tax applies.

The funds collected by the IRS and managed by the DOL would be directed toward a set of targeted measurements designed to

support displaced workers. First, resources will be allocated towards community colleges to develop AI literacy programs. Next, additional revenue will go towards establishing employee retention and reskilling grants, which will partially offset contribution obligations for firms that elect to retrain workers rather than conduct layoffs. Finally, remaining funds will be allocated towards accounts with national lifelong learning credit accounts modeled after programs like Singapore's.

Year 3: Pilot Phase

The pilot tax will begin with a three year pilot program in order to empirically evaluate and minimize implementation risks. Limiting the initial scope allows policymakers to refine the tax's definitions, compliance standards, and funding mechanisms before initiating a broader rollout. It also enables greater data collection on labor outcomes and firm behavior. These two factors are essential for assessing whether or not the robot tax meaningfully improves worker transitions.

Years 4-5: Evaluation and Sunset

At the conclusion of the three year period, the policy would be subject to a formal evaluation from the GAO. This review will assess the program's effectiveness based on the metrics established by the DOC. These metrics will include reemployment rates for displaced workers, participation rates in retraining programs, and employer compliance levels. Including a mandatory review makes the tax's continuation based on measurable outcomes and reassures stakeholders that the program is evidenced based rather than political.

Year 6+: Scaling

If the GAO finds that more than fifty percent of displaced workers are successfully employed within one year, the policy will trigger an expansion clause. The program can then begin gradually expanding reskilling credits and retention subsidies as administrative systems mature. Any further expansion would include an additional five year sunset provision to ensure continued congressional oversight and reassessment.

Conclusion

AI will undoubtedly transform how we work, but a robot tax can help equalize who benefits from said transformation. The AI era posits unique challenges for the labor market caused by adjustment lags, firm externalities and skill mismatches. A robot tax can address all these concerns by internalizing costs, ensure long term competitiveness, and secure geopolitical leadership.

Rather than hindering the free market, a robot tax enhances its adaptive capabilities. Using the pilot program structure will make the robot tax flexible, measurable, and resilient enough to transform automation into opportunity rather than disaster. With a robot tax, the U.S. does not have to choose between innovating AI technologies and workforce stability. Instead, we can make sure that the next Industrial Revolution works for everyone.

References

- Acemoglu, Daron, and Pascual Restrepo. "Artificial Intelligence, Automation and Work." NBER, January 15, 2018. <https://www.nber.org/papers/w24196>.
- Benhamou, Salima. "Artificial Intelligence and the Future of Work." *Revue d'économie industrielle*, January 5, 2023. <https://journals.openedition.org/rei/8727#>:
- Berg, Janine. "Minimizing the Negative Effects of Ai-Induced Technological Unemployment." International Labour Organization, September 18, 2025. <https://www.ilo.org/resource/article/minimizing-negative-effects-ai-induced-technological-unemployment>.
- Cazzaniga, Mauro, Florence Jaumotte, Longji Li, Giovanni Melina, Augustus J Panton, Carlo Pizzinelli, Emma J Rockall, and Marina Mendes Tavares. "Gen-AI: Artificial Intelligence and the Future of Work." Staff Discussion Notes 2024, no. 001 (January 14, 2024): 1. <https://doi.org/10.5089/9798400262548.006>.
- Cook, William. "Automation Exposure Score." LMI Institute, November 7, 2024. <https://www.lmiontheweb.org/automation-exposure-score/>.
- Drydakakis, Nick. "Artificial Intelligence and Labor Market Outcomes." IZA World of Labor, February 12, 2025. <https://wol.iza.org/articles/artificial-intelligence-and-labor-market-outcomes/long>.
- Ebbin, Adam P. "Letter to VA Lawmakers: Oppose Overbroad AI Regulation Creating Excessive Disclosures, Strict Liability, and Vague Enforcement Standards (SB 365)." Chamber of Progress, February 4, 2026. <https://progresschamber.org/resources/letter-to-va-lawmakers-oppose-overbroad-ai-regulation-creating-excessive-disclosures-strict-liability-and-vague-enforcement-standards-sb-365/>.
- Frazier, Kevin T. "AI Is Transforming the Economy—Not Destroying It." Cato.org, January 27, 2026. <https://www.cato.org/blog/ai-transforming-economy-not-destroying-it>.
- Gallant, Erik. "Automation and a Changing Economy: The Case for Action." Aspen Institute, May 23, 2025. <https://www.aspeninstitute.org/publications/automation-and-a-changing-economy-the-case-for-action/>.
- Hatzius, Jen, Joseph Briggs, Devesh Kodnani, and Giovanni Pierdomenico. "The Potentially Large Effects of Artificial Intelligence on Economic Growth (Briggs/Kodnani)." Goldman Sachs Publishing, March 26, 2023. <https://www.gspublishing.com/content/research/en/reports/2023/03/27/d64e052b-0f6e-45d7-967b-d7be35fabd16.html>.
- Heer, Carsten. "AI In Robotics - New Position Paper." International Federation of Robotics, February 10, 2026. <https://ifr.org/ifr-press-releases/news/ai-in-robotics-new-position-paper>.
- ILO. "Artificial Intelligence Adoption and Its Impact on Jobs." International Labour Organization, August 28, 2025. <https://www.ilo.org/publications/artificial-intelligence-adoption-and-its-impact-jobs>.
- Jacinto, Nuno. "RoboTax." NOVA Tax Research Lab, November 29, 2023. https://taxlab.novalaw.unl.pt/?page_id=4349.
- Kamal-Chaoui, Lamia. "Place-Based Policies for Displaced Workers." OECD Local Economic and Employment Development (LEED) Papers, June 30, 2025. https://www.oecd.org/en/publications/place-based-policies-for-displaced-workers_69b86784-en.html.
- Kratsios, Michael J, David O Sacks, and Marco A Rubio. "Winning the Race: America's AI Action Plan." WhiteHouse.gov, July 2025. <https://www.whitehouse.gov/wp-content/uploads/2025/07/Americas-AI-Action-Plan.pdf>.
- Lin, Chenghua, and Yanyi Wu. "The Hidden Multiplier: Unraveling the True Cost of the Global AI Skills Gap." IEEE Technology and Society, February 28, 2025. <https://technologyandsociety.org/the-hidden-multiplier-unraveling-the-true-cost-of-the-global-ai-skills-gap/>.

Tr

Mann, Roberta F. "I Robot: U Tax? Considering the Tax Policy Implications of Automation." *McGill Law Journal*, June 1, 2019. <https://lawjournal.mcgill.ca/article/i-robot-u-tax-considering-the-tax-policy-implications-of-automation/>.

Metzenbaum, Howard. "Worker Adjustment and Retraining Notification Act." *GovTrack*, August 4, 1988. https://www.govtrack.us/congress/bills/100/s2527/text#-google_vignette.

Randolph, Liane M. "2025 ANNUAL REPORT Cap-and-Trade Auction Proceeds." *California Air Resources Board*, May 2025. https://ww2.arb.ca.gov/sites/default/files/auction-proceeds/cci_annual_report_2025.pdf.

Samman, Daniel, Ekkehard Ernst, Miguel Sanchez Martinez, Richard Home, and Stefan Kühn. "World Employment and Social Outlook Trends 2023." *ILO Flagship Report*, November 6, 2022. https://www.ilo.org/sites/default/files/wcmsp5/groups/public/@dgreports/@inst/documents/publication/wcms_865332.pdf.

West, Darrell M., Aylin Caliskan Kyra Wilson, Monica Anderson Nicol Turner Lee, and William A. Galston. "What the Public Thinks about AI and the Implications for Governance." *Brookings*, April 17, 2025. <https://www.brookings.edu/articles/what-the-public-thinks-about-ai-and-the-implications-for-governance/>.

Zakrzewska, Ada. "OECD Employment Outlook 2025." *OECD*, July 9, 2025. https://www.oecd.org/en/publications/2025/07/oecd-employment-outlook-2025_5345f034/full-report/staying-in-the-game-skills-and-jobs-of-older-workers-in-a-changing-labour-market_cc7ee11c.html.

THIS PAGE LEFT INTENTIONALLY BLANK.

The Virtuous Insider

Why Prediction Markets Need the Traders We Usually Ban

Amr Yakout¹

Introduction

Controversy has surrounded prediction markets – such as Polymarket and Kalshi – since their rise in popularity and penetration of social and cultural trends. A growing number of users have dedicated significant effort to investigating insider accounts, and policy conversations have emerged around whether insider activity constitutes ‘illegal’ insider trading. Representative Ritchie Torres of New York introduced the Public Integrity in Financial Prediction Markets Act of 2026 in response to suspicious trading activity on these platforms (Torres 2026). Most experts agree there is significant insider trading on these platforms, but the question remains: is this a failure to be regulated, or a foundational mechanism to the success of these markets?

Conventional economic and legal wisdom suggests that insider trading is detrimental to markets and society. That makes sense in the stock market, as it destroys trust, is unfair, and discourages corporate investment (Manove 1989). Capital gain based on asymmetric information in a market where the primary goal is to raise capital (i.e., the stock market) is simply cheating. However, what if the goal behind prediction markets isn’t to raise capital, but rather information discovery? I argue that, while the good produced by a stock market is capital, the good produced by a prediction market is information. As such, we should allow ‘insider trading’ to maximize allocative efficiency in the production of information and discourage market regulation that eliminates the insider’s ability to participate. After all, the insiders have welfare enhancing effects when it comes to price discovery and information production.

Argument and Analysis

A prediction market would not serve as a good reflection of public opinion, nor as a good ‘predictor’, if it forecast a 90% chance of a candidate winning but the candidate lost. Accuracy of information is key to the success of these markets and is fundamental to their survival. The morality and politics surrounding these markets, and the ethics around making money from them, are not the concern of this article, though I will briefly comment on why these markets exist and their utility.

To understand the utility of insiders to prediction markets, I must start by explaining what prediction markets are, how they differ from stock markets, and what utility they provide to markets and society. For the sake of this article, I am restricting my analysis to contract-based prediction markets – Kalshi and Polymarket. Both function as winner-take-all (or winner-takes-most) markets where traders bet on event outcomes, paying out the correct side. We can also think of them as a form of options trading, but with a much more intuitive and accessible framework. Both platforms are synthesizing crowd wisdom into probabilities for future events like elections or economic shifts. But what utility do these platforms provide? The literature on this long predates Kalshi and Polymarket, largely because sports betting, insurance, and future contracts are all forms of prediction markets to a certain extent. For example, Wolfers and Zitzewitz (2004) argue that “the power of prediction markets derives from the fact that they provide incentives for truthful revelation, they provide incentives for research and information discovery, and the market provides an algorithm for aggregating opinions.” Fundamentally, prediction markets function as ‘truth engines’ by solving the problem of aggregated knowledge.

The Hayekian argument in his seminal 1945 paper, *The Use of Knowledge in Society*,

1 Amr Yakout is an MPP student at the McCourt School of Public Policy at Georgetown University

applies here: information is not concentrated in a single mind but is dispersed as incomplete bits of knowledge among many individuals. For example, a factory floor worker creates asymmetric information when they discover a shipment is late before the CEO does. A poll worker sees turnout trends before the pundit, or even candidate. By incentivizing these agents to trade on their private information, prediction markets bypass hierarchy and instantly incorporate that data into the price. The market pays them to tell the truth.

There is another argument for the utility of prediction markets to society: they provide a hedging opportunity to other market players, a form of insurance. These markets allow businesses to neutralize risks that traditional insurance companies won't cover, like election outcomes, Federal Reserve rate increases, or specific regulatory approvals. Kalshi actually markets itself with that value proposition on their 'about' page.

Think of a change in administrations that you believe will have downstream effects in terms of tariffs, resulting in \$1 million in additional costs to a small business. As a hedge, the owner can bet heavily on the administration to win. If it wins, you lose \$1 million in tariff costs but make up some by winning the bet. If the administration loses, then you aren't liable for that extra \$1 million.

If you accept that the value of these markets is maximal truth-seeking, which allows for accurate information discovery and better hedging against risk, then our argument for insider trading becomes stronger. If the entire purpose of the market is to give society an accurate early warning (e.g., "Will this vaccine fail clinical trials?"), then banning the scientist who knows the vaccine is failing is counterproductive. We are sacrificing the societal value of the truth (warning investors/patients early) to preserve the optical value of fairness (making sure the scientist doesn't make a quick buck). But do we know that insider traders improve information discovery?

To understand why insider trading improves information discovery in this instance, we must revisit George Akerlof's "The Market for Lemons" (1970). Akerlof famously demonstrated how asymmetric information can destroy a market: if buyers cannot distinguish

between high-quality cars and defective ones ("lemons"), they will assume every car is a lemon, eventually driving honest sellers out of the market. This process is known as adverse selection.

In a prediction market, however, we can invert this logic to our advantage. By allowing insiders to trade, we intentionally introduce a severe information asymmetry. By definition, an insider possesses a set of private information $I_{Private}$ regarding the event outcome that is superior to the public information set I_{Public} available to the crowd. Traditional equity markets view this imbalance as unfair, especially as it crowds out investment from other market players who need assurances they are not trading against insiders. However, prediction markets view this imbalance as untapped utility. The goal is not to protect the outsider from the insider, but to incentivize the insider to leak their private information to the public.

Uncertain predictors, or traders who are merely guessing or betting on intuition, realize they are possibly trading against individuals who know the outcome with certainty. Combined with the transparency that comes with prediction markets in the form of betting and outcome history, reasonably rational outsiders can identify the risk of an insider being there and any perceived disadvantage that comes with that. Facing this perceived disadvantage, rational "noise traders" withdraw, much like Akerlof's buyers. While standard economic theory views this "crowding out" as a failure of liquidity, in the context of a truth engine, it is a success of filtration. By driving out the guessers or diluting their influence on the market, we can minimize noise from the price discovery mechanism as much as possible, leaving behind only the signal provided by the informed. The resulting price may have lower trading volume, but it possesses significantly higher informational fidelity.

One might argue that this 'purification' of the market leads to a liquidity death spiral in the form of lower trading volume. If the reverse-Akerlof logic holds and 'noise traders' flee, we are left with a thin market of insiders trading against one another. However, this critique mistakenly views the prediction market as a zero-sum casino rather than a mechanism for information extraction. The solution to the liquidity trap lies in the Automated Market Maker (AMM) (Semrau 2025). In modern

frameworks – like the Logarithmic Market Scoring Rule used by prediction platforms – the market organizer does not need a counterparty for every trade. The AMM is algorithmically programmed to accept trades at slippage-adjusted prices, effectively acting as a ‘buyer of last resort’ for information. The AMM is designed to lose money to informed traders as a cost of purchasing the truth. By subsidizing the liquidity, the AMM ensures that even if only one insider exists in the world, they have an incentive to put money where their mouth is.

Critiques and Proposal

A common critique is that admitting insiders introduces moral hazard: a goalkeeper might bet against their own team and intentionally miss a save, creating an outcome driven by the market rather than the reverse. Furthermore, specific in-game events, such as a goalkeeper error, are too common to distinguish from intentional manipulation, thereby validating the moral hazard concern.

The standard economic counter-argument to this critique focuses on the market mechanism itself. In microeconomic terms, the ‘forensic alarm’ is not triggered by the event itself, but by the marginal change in order flow preceding it. A rational agent (the goalkeeper) faces a strict dilemma defined by liquidity constraints, where if they bet on a highly probable loss to minimize suspicion, the market is liquid. But the return on investment is negligible for such a bet, failing to outweigh the immense opportunity cost of a career ban. Conversely, if they attempt to maximize profit by betting on an unlikely upset, they face a thin market where any substantial wager causes massive price slippage. This creates a paradox for the fixer: the only trades profitable enough to justify the risk are the exact trades that generate statistically anomalous capital concentrations. Thus, the market does not need to distinguish between a clumsy save and a fixed one; it simply needs to identify the ‘insider’ money that moved before the mistake happened, effectively rendering the price mechanism a deterrent rather than an incentive.

However, this deterrence logic assumes several conditions that may not always hold. It presumes rational actors with long time horizons, but what about one-off situations such as a retiring athlete, someone facing financial desperation, or actors who underestimate

detection risk? The argument also assumes effective detection mechanisms exist, but sophisticated actors can obscure trades through intermediaries or by fragmenting positions across multiple accounts. To this end, distinguishing types of insider activity is important.

The distinction between “passive” insiders (who possess information but cannot influence outcomes) and “active” insiders (who can manipulate outcomes directly) warrants special policy treatment. A CDC scientist knowing vaccine trial results represents pure information asymmetry—beneficial under this framework. A goalkeeper, or a political actor, represents potential manipulation risk. The line between these categories, however, can blur in practice.

Consider the recent controversy surrounding Karoline Leavitt, the current White House Press Secretary, who ended a press conference approximately 30 seconds before the 65-minute threshold that was simultaneously trading on Kalshi. The market had priced “Yes” (over 65 minutes) at 98% probability, yet “No” bettors made nearly 49x returns when she abruptly stopped the briefing at 64:33. While we cannot determine whether this was coincidental, the incident illustrates the difficulty of distinguishing between an insider with advance knowledge of an outcome and an actor with the power to determine that outcome. Whether prediction markets can or should enforce such distinctions, and whether they can effectively detect when outcome-controllers are participating, remains an open question and a critical avenue for regulation.

Mitigating these issues is a central public policy question that needs to be addressed prior to removing insider trading restrictions on prediction markets. The answer this analysis offers is that insider trading on prediction markets is not a monolithic phenomenon. The important distinction is between traders who possess superior information about an outcome and traders who possess the ability to determine that outcome.

Consider two traders on a prediction market contract asking “Will Drug X pass FDA clinical trials?” The first is a research scientist at the pharmaceutical company running the trial. She has seen preliminary data suggesting the drug will fail, and she trades accordingly.

She possesses asymmetric information, but she cannot change the trial's outcome; the data are what the data are. The second trader is a senior FDA official with direct authority over the approval decision. He is not 'merely' informed about the likely outcome because he can alter it. Both traders profit from private knowledge, but only one of them is trading on an event they can also control. The scientist improves the market's informational accuracy by injecting a true signal, while the FDA official corrupts it by introducing the possibility that the outcome was manufactured to match his position. Prediction markets should permit the former and prohibit the latter. The operative question for any given trader is not "do you know something others don't?" but rather "can you, or can someone in your network, change the outcome?" Insider trading should be allowed only when the answer to the second question is no.

Enforcing this distinction requires an enhanced identity verification process in which prediction markets classify users according to their capacity to influence outcomes in specific event categories. Standard Know Your Customer (KYC) confirms that a user is who they claim to be. This 'Authoritative' KYC goes further, requiring users to disclose roles or affiliations that grant them outcome-altering power. A government official opening an account, for example, would be flagged as a "designated user" in political event categories. Their trades in those categories would be subject to automated monitoring, and the platform would be required to flag and report anomalous activity to the relevant regulatory body.

This does not mean all designated users are banned from trading. A Congressional staffer could still trade on "Will the Fed raise rates?" because they have no role in executive monetary policy. The designation is category-specific, not a blanket restriction. The platform's obligation is threefold: identify users with potential outcome authority in defined categories, monitor their trading activity in those categories with heightened scrutiny, and report patterns consistent with outcome manipulation to regulators.

Conclusion

The goal of this article was to identify one of the key perceived weaknesses of prediction markets, the lack of proper insider trading

stipulations, and argue that it's actually one of its biggest strengths. In arguing that, I discuss the utility of prediction markets to society in the hopes that I counter some of the pessimistic narratives arising from policy professionals and others regarding these markets. Using classic economic concepts such as adverse selection, moral hazard, and asymmetric information, I have framed this through the lens of economic theory, though I leave an avenue for further advancements in studying the unintended consequences of these markets, especially their explosive penetration of culture and social media. This is not an argument condoning gambling or participation in prediction markets, but I am outlining the counterintuitive benefits to society associated with these controversial practices.

References

- Akerlof, George A. 1970. "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism." *Quarterly Journal of Economics* 84 (3): 488–500.
- Kim, Oihyun. 2026. "White House Briefing Fuels Insider Trading Debate—Pelosi Joins the Irony." *BeInCrypto*, January 11, 2026. <https://beincrypto.com/white-house-briefing-insider-trading-debate/>.
- Corporate Finance Institute. 2023. "Slippage - Definition, Why It Happens, How To Minimize." Corporate Finance Institute, October 19, 2023. <https://corporatefinanceinstitute.com/resources/career-map/sell-side/capital-markets/slippage/>.
- Grantham-Philips, Wyatt. 2026. "A \$400,000 Payout After Maduro's Capture Is Putting Prediction Markets in the Spotlight." *Associated Press*, January 11, 2026. <https://www.pbs.org/newshour/nation/a-400000-payout-after-maduros-capture-put-prediction-markets-in-the-spotlight-heres-how-they-work>.
- Hayek, F. A. 1945. "The Use of Knowledge in Society." *American Economic Review* 35 (4): 519–30.
- Manove, Michael. 1989. "The Harm From Insider Trading and Informed Speculation." *Quarterly Journal of Economics* 104 (4): 823–45.
- Semrau, Jan Daniel. 2025. "Automated Market Making on Kalshi: How Cognitive Financial Agents Optimize Stochastic Control Problems." *Encyclopedia Autonomica* (blog), February 19, 2025. <https://jdsemrau.substack.com/p/automated-market-making-on-kalshi>.
- Torres, Ritchie. 2026. "In Response to Suspicious Polymarket Trade Preceding Maduro Operation, Rep. Ritchie Torres Introduces Legislation to Crack Down on Insider Trading on Prediction Markets." Press release, January 9, 2026. <https://ritchietorres.house.gov/posts/in-response-to-suspicious-polymarket-trade-preceding-maduro-operation-rep-ritchie-torres-introduces-legislation-to-crack-down-on-insider-trading-on-prediction-markets>.
- Wolfers, Justin, and Eric Zitzewitz. 2004. "Prediction Markets." *Journal of Economic Perspectives* 18 (2): 107–26. fund human capital (Gallant 2025). Using this framework, the robot tax ensures that innovation remains economically, efficient, politically legitimate, and socially sustainable.

THIS PAGE LEFT INTENTIONALLY BLANK.

The Nearshoring Prescription

A North American Pharmaceutical Chapter

Lynzee Lee¹

Abstract

The United States-Mexico-Canada Agreement (USMCA) facilitates trade in goods across North America but does not sufficiently address the regulatory complexity of pharmaceuticals. The agreement lacks a coordinated framework for pharmaceutical regulation, limiting supply chain resilience and timely access to medicines. This memo recommends that Mexico proposes adding a dedicated Pharmaceutical Chapter to the USMCA during the July 2026 Review to establish a framework for trilateral cooperation aimed at attracting pharmaceutical manufacturing to Mexico, strengthening North American pharmaceutical supply chains, and streamlining regulatory coordination and approvals. The proposal aligns with Mexico's nearshoring strategy and the Trump administration's efforts to reshore supply chains and reduce prescription drug prices, creating mutual benefits for all parties. Pharmaceuticals merit a standalone chapter because, unlike most traded goods, they are subject to complex regulatory approval processes and are closely linked to public health and national security. The proposed chapter would enable coordinated regulatory cooperation among Mexico, the United States, and Canada, thereby strengthening regional supply chains while improving the access and affordability of medicines across North America.

Background of Free Trade Agreements in North America

The Canada-United States Free Trade Agreement (CUSFTA) went into effect January 1, 1989. As one of the first comprehensive bilateral free trade agreements, this agreement between Canada and the United States “included the elimination of tariffs, the reduction of many non-tariff barriers, and it was among the first trade agreements to address trade in services” (Global Affairs Canada 2021). It represented an important shift towards free trade on the continent. Subsequently, in the early 1990s, Mexico, Canada, and the United States began their formal negotiations to expand the CUSFTA to also include Mexico, as requested by the president of Mexico at the time, Carlos Salinas de Gortari.

In December 1992, NAFTA was signed by Mexican President Carlos Salinas de Gortari, Canadian Prime Minister Brian Mulroney, and US President George H.W. Bush; it officially took effect on January 1, 1994. The main purpose of NAFTA was to establish a free trade area between Mexico, Canada, and the United States. In current US dollars, the combined gross domestic product (GDP) of all three countries in 1994 was \$8.42 trillion (World Bank 2025). As a regional trading bloc of this magnitude, it was believed that all three countries could

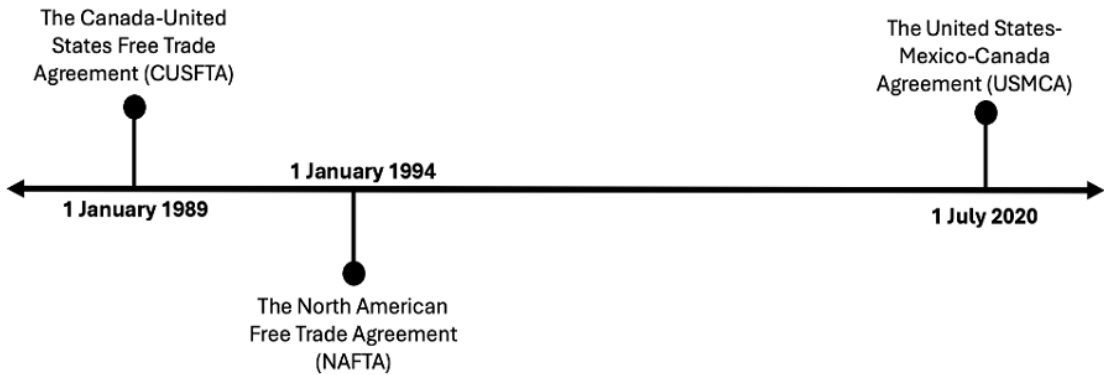
“benefit from reduced costs, more competitive prices, and greater global trading power” (Aguilar 1992). NAFTA's objectives, as stated in the official 1992 document, include eliminating trade barriers of both goods and services, increasing fair competition, expanding investment opportunities, protecting intellectual property rights, implementing procedures to enforce NAFTA and settle disputes, and expanding multilateral cooperation.

Arguably the most fundamental component of the document was a gradual elimination of tariffs on most goods and services over a fifteen-year period. This aimed to make North American goods more competitive globally by lowering costs. Another crucial piece of the document was the “rules of origin,” which help determine whether goods are eligible to be tariff-free, helping ensure that NAFTA's trade benefits are solely for products truly originating in North America and not simply passed through a NAFTA country after having been manufactured abroad (NAFTA 1992).

NAFTA was primarily designed to eliminate tariffs, reduce regulatory barriers, and protect intellectual property, focusing on broad economic liberalization rather than sector-specific strategies. As a result, pharmaceuticals were treated largely as standard manufactured goods, without tailored provisions to address

1 Lynzee Lee is an MIDP student at the McCourt School of Public Policy at Georgetown University

Timeline of free trade agreements in North America



their unique regulatory complexity. NAFTA did not establish a dedicated framework for pharmaceutical regulatory coordination, supply-chain resilience, or emergency preparedness, areas that are typically only partially addressed, if at all, in most existing trade agreements. While the USMCA later modernized trade rules to reflect evolving economic priorities, it did not fundamentally change the treatment of pharmaceuticals or introduce a comprehensive, sector-specific regulatory framework.

Calls to renegotiate NAFTA emerged during the Obama administration, primarily framed around the need to modernize the agreement to reflect 21st-century economic realities, including digital trade, labor, and environmental standards. These motivations shifted during the 2016 US presidential race, when both Donald Trump and Senator Bernie Sanders criticized NAFTA for contributing to job losses and manufacturing decline in the United States (Chatzky, McBride, & Sergie 2020). After taking office, President Trump made renegotiation a central priority, pursuing a more protectionist and domestically focused agenda aimed at rebalancing trade relationships and strengthening US manufacturing, ultimately resulting in the replacement of NAFTA with the USMCA on July 1, 2020. The core framework of NAFTA remains the same for USMCA, but a few key differences exist. According to Lobosco, Fung, and Luhby (2019), these include:

- Automobile manufacturing rules: raised the required percentage of vehicle parts made in North America
- Increased minimum wages for certain workers

- Labor regulation enforcement: introduced new enforceable labor standards
- An increase in environmental protections
- Removed previous restrictions on Canadian dairy and poultry market: increased the amount of dairy and poultry the US can export to Canada

Added new benefits for US tech industry: prevented Canada and Mexico from requiring US companies to store data locally and protect them from legal liability for certain internet content

During the USMCA negotiations, pharmaceutical provisions were expanded significantly, due in part to lobbying efforts by the US pharmaceutical industry (Palmer 2019; Kirchner & Rieras 2020). Compared to NAFTA's minimal treatment, where the word "pharmaceutical" appeared only five times, the USMCA includes over 100 references to the word and introduces Annex 12-F on Pharmaceuticals. This annex builds on broader USMCA commitments related to regulatory coherence and technical barriers to trade, applying them specifically to the pharmaceutical sector. It establishes common definitions, outlines regulatory procedures, reduces duplicative requirements, encourages international harmonization, and strengthens cooperation in inspections and information sharing.

While these provisions represent a step forward, they remain focused on regulatory processes rather than establishing a fully integrated framework for supply-chain resilience or emergency preparedness. While the United States primarily led pharmaceutical

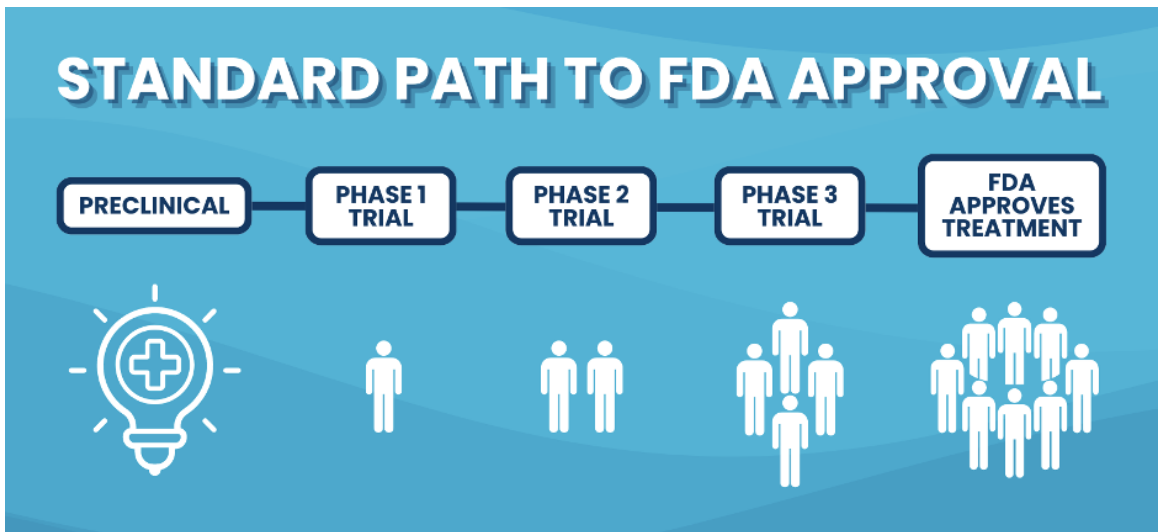


Figure 1: Standard Path to FDA Approval

Note: Figure from Kunde, Juhi. (2024, August 27). How do drugs get approved (and fast-tracked) by the FDA? LUNgevity.

changes in the original USMCA negotiations, the current policy environment—particularly in light of the agreement’s upcoming July 2026 review—creates an opportunity for Mexico to take a more proactive role. Although the review is expected to focus largely on the Trump administration’s short-term priorities, such as migration and drug trafficking, it also provides a strategic opportunity for member countries to propose reforms that strengthen North America’s broader economic integration. In this context, Mexico is well positioned to put forward a new pharmaceutical chapter that would expand its role as a nearshoring hub for pharmaceutical production.

Analysis: Current Regulatory Barriers and Inefficiencies

Pharmaceuticals warrant their own chapter in the USMCA because they are fundamentally different from typical manufactured goods, being heavily regulated and subject to lengthy approval processes. In the United States, the US Food and Drug Administration (FDA), the body overseeing pharmaceuticals, asserts that “American consumers benefit from having access to the safest and most advanced pharmaceutical system in the world” (US Food and Drug Administration 2022), with the approval process for a new drug averaging 12 to 15 years (Kunde 2024). This process begins with laboratory and animal testing, followed by multiple clinical trials in humans to assess

safety and efficacy; FDA reviewers generally require evidence from at least two well-designed trials, although one may suffice for rare or life-threatening diseases. Approval indicates that a drug’s benefits outweigh its risks based on clinical data, disease severity, and available alternatives, and all approved drugs must carry FDA-approved labels communicating proper use, benefits, and risks. Regulatory oversight continues throughout a drug’s lifecycle through postmarketing surveillance and batch-level quality controls (US Food and Drug Administration 2020), and is embedded in every step of manufacturing for both active pharmaceutical ingredients (APIs) and finished medicines. Both domestic and international factories are required to follow these rules and are subject to inspection by the FDA (Code of Federal Regulations 2009).

Mexico and Canada have similar regulatory agencies, Comisión Federal para la Protección contra Riesgos Sanitarios (COFEPRIS) and Canada’s Drug Agency (CDA), respectively. Mexico’s COFEPRIS is similar to the FDA and oversees the manufacturing, import, and export of pharmaceuticals in Mexico. Founded in 2001 under the General Health Law, it is led by a commissioner appointed by the president. Differences between the FDA and COFEPRIS regulatory systems can create challenges for pharmaceutical companies, including duplicate testing, longer approval timelines, and higher costs. Exporters must

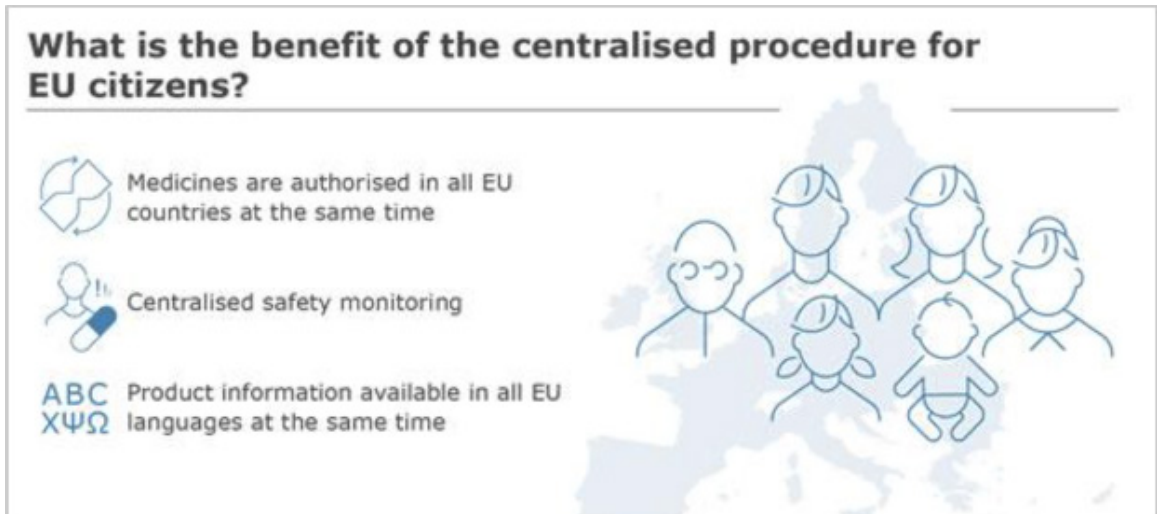


Figure 2: Benefits of a centralized pharmaceutical regulatory authority

Note: Figure from European Medicines Agency. (2024, July 31). Authorisation of medicines.

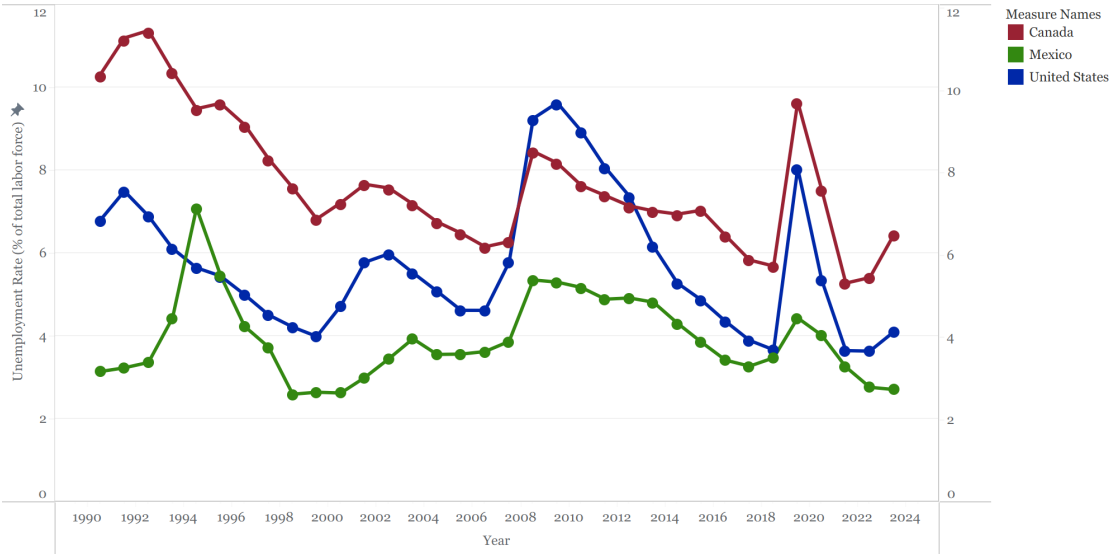
frequently submit similar safety and efficacy data to both agencies, slowing the introduction of new medicines and increasing compliance expenses. Because pharmaceutical products are subject to such extensive regulatory review, differences in approval processes, standards, and timelines across countries can act as significant trade barriers, underscoring the need for a dedicated pharmaceutical chapter in the USMCA that promotes regulatory alignment while maintaining high safety and quality standards across all three countries. One way to envision a coordinated approach is by looking at the European Union, where the European Medicines Agency (EMA) serves as the centralized authority overseeing pharmaceutical regulation across member states.

The European Union (EU) is a political and economic union of 27 member states that share common institutions, policies, and regulations to facilitate trade, economic integration, migration, and cooperation. Unlike the USMCA, which is a free-trade agreement between three countries, the EU functions more as a centralized regulatory and political bloc. The EU and USMCA differ in many ways, most notably in legal authority, market integration, regulatory streamlining, and how they treat external countries: EU members apply a common external tariff to non-members, while USMCA members have independent external tariff policies for non-members. Despite these differences, both frameworks share the main goal of reducing trade barriers and promoting

economic cooperation, and thus the EU's European Medicines Agency (EMA) is a useful comparative model for how a more coordinated and centralized pharmaceutical regulatory framework could reduce inefficiencies and strengthen regional supply-chain resilience under the USMCA.

The EU's European Medicines Agency (EMA) participates in the following activities to reduce pharmaceutical trade barriers and inefficiencies (European Medicines Agency 2024):

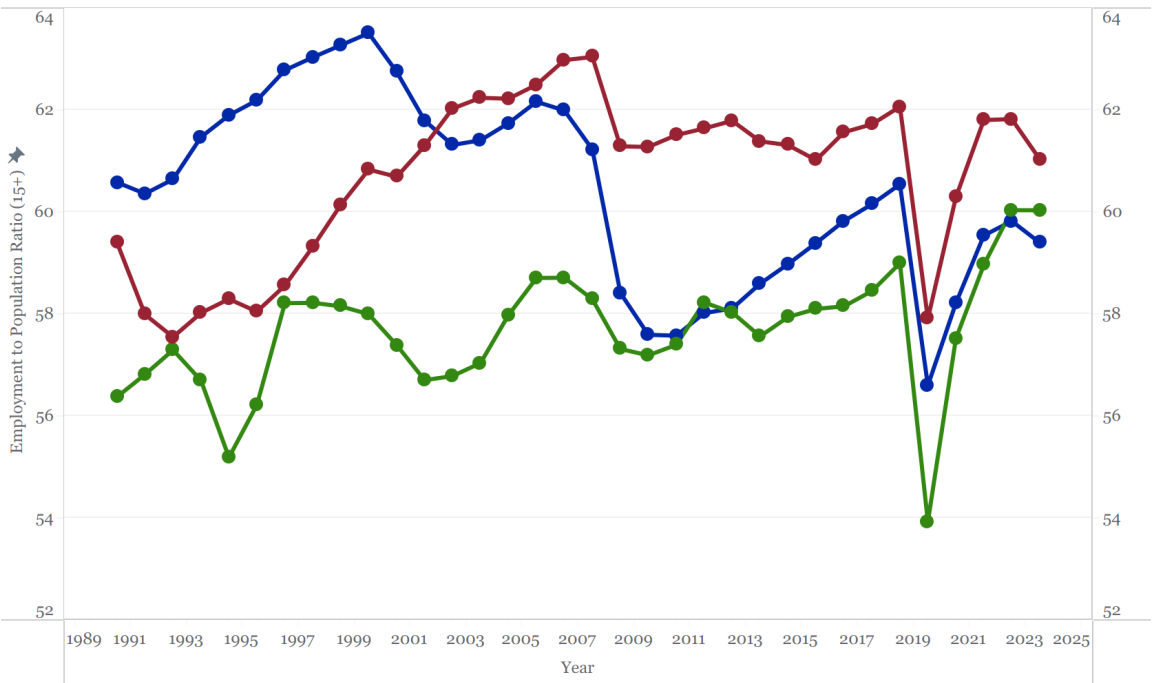
- Operates under EU pharmaceutical legislation, which provides the legal framework for how medicines are developed, assessed, authorized, and monitored across all EU member states
- Reviews new medicines to ensure they are safe and effective, using a centralized procedure that allows companies to submit clinical trial data, which EMA then evaluates and once approved allows a medicine to be marketed and become available across all EU countries simultaneously
- Conducts research to improve the evaluation, authorization, and safety monitoring of medicines, ensuring that scientific evidence is the basis for its regulatory decisions
- Ensures preparedness for health crises by providing guidance, scientific advice, and coordinated actions to ensure the



Source: World Development Indicators, 2025.

Figure 3: Unemployment total as a percentage of total labor force for Mexico, Canada, and USA

Note: Author’s calculation based on data from World Bank (2025)



Source: World Development Indicators, 2025.

Figure 4: Employment to population ratio (15+) for Mexico, Canada, and USA

Note: Author’s calculation based on data from World Bank (2025)

availability of medicines during public health emergencies

One of the most important components of EMA is its preparedness and cooperation strategy during a public health emergency. EMA engages in the following activities to prepare for crises:

- Tracks medicine supply and reserves across EU countries using networks built by EMA and the EU, allowing them to spot potential shortages ahead of time
- Coordinates emergency responses by sharing information, aligning actions, and advising on safety, quality, and efficacy of medicines during emergencies
- Runs an emergency task force to review evidence, support clinical trials, and use real-world data to speed up medicine approvals and monitoring during public health emergencies
- Communicates with health experts, government officials, and the public in a clear and transparent manner, allowing all stakeholders to stay informed during emergencies

This health crisis plan enabled EMA to activate its emergency response during the COVID-19 pandemic, allowing for a streamlined response during a critical time. EMA took part in the following during the pandemic:

- Used rolling reviews and conditional marketing authorizations to evaluate vaccines and treatments as data became available, shortening typical approval timelines
- Tracked safety in real time, issued updates, and adjusted guidance
- Quickly approved new vaccine manufacturing sites and raw material suppliers to increase supply of new COVID-19 vaccines
- Released assessment reports promptly and kept the public and stakeholders informed to maintain trust during fast-moving regulatory decisions
- Used its medicine tracking technology to collect up to date information on shortages of essential drugs

In the EU, each member state still has its own regulatory body for domestic pharmaceutical production who works in cooperation with EMA to conduct audits and inspections. This model illustrates how complete regulatory harmonization is not necessary, but a centralized regulatory authority that can streamline approvals, reduce duplication, and coordinate emergency responses across multiple countries is crucial. The proposed new chapter of the USMCA should provide information on the creation of a governing body similar to the EMA for North America that would work in conjunction with the FDA, COFEPRIS, and CDA.

Analysis: Why Mexico Would Benefit

Increasing pharmaceutical production in Mexico aligns with many of Mexico's stated goals, including its effort to expand nearshoring and decrease labor informality. About two years ago, the former president of Mexico, Andrés Manuel López Obrador (AMLO), issued a decree that Mexico would provide incentives to expand nearshoring in key industries, including pharmaceutical manufacturing (UNCTAD 2023). This presents a great opportunity for Mexico to boost jobs in this sector, at a time when the Trump administration wants to reduce the cost of medicines in the United States.

Reducing labor informality is also a major policy goal for Mexico because informal employment limits tax revenue, reduces social insurance coverage, and often provides lower wages and fewer protections for workers (MBN Staff 2025). By expanding nearshoring in high-value industries like pharmaceuticals, Mexico could create more formal, registered jobs that would help workers gain benefits and allow the government to collect more taxes. Nearshoring also encourages investment in domestic production facilities, which can absorb more of Mexico's working-age population into formal employment and support economic stability while meeting the growing demand for pharmaceutical products in North America.

Labor market data in Mexico highlight the opportunity to create incentives for more formal employment. As shown in Figure 3, Mexico has the lowest unemployment rate among the USMCA members, yet it also has the lowest employment-to-population ratio (see Figure 4),

Exports

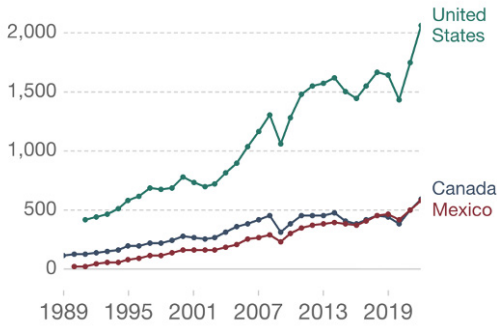


Figure 5 (left): Total value of all exports in USD billions for Mexico, Canada, and USA

Imports

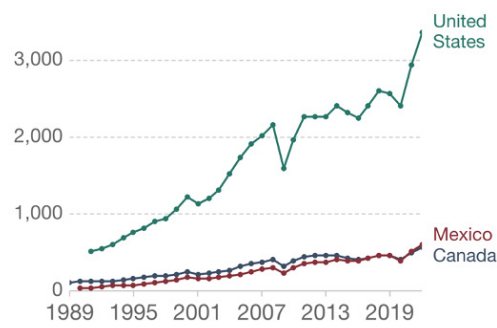


Figure 6 (right): Total value of all imports in USD billions for Mexico, Canada, and USA

Note: Data and graph retrieved from Global LAC Datahub (2025) <https://globallacdatahub.com/country-level-trade.html>

suggesting that many working-age adults are in informal or irregular employment. Increasing the number of formal pharmaceutical manufacturing jobs through nearshoring could help raise the employment-to-population ratio, absorb more of the working-age population into registered employment, and align with Mexico’s broader goals of labor formalization.

Over the past several decades, US exports and imports have skyrocketed relative to Mexico and Canada (see Figures 5 and 6). This does not mean that Mexico’s and Canada’s trade

have not grown, but rather both countries have seen more gradual and steady growth. Mexico can benefit from targeting high-value, strategic sectors like pharmaceuticals to accelerate export growth. The proposed Pharmaceutical Chapter in the USMCA could help create procedures to establish a more efficient and cooperative North American pharmaceutical supply chain, allowing Mexico to move up the value chain and capture more exports.

The pharmaceutical industry in Mexico has grown in the recent decades and in 2023, “Mexico exported 165 of the 350 pharmaceutical and [active pharmaceutical ingredient] products designated as critical by the International Trade Administration to the United States” (Kucik & Rudman 2024). Mexico already has existing pharmaceutical manufacturing facilities; Figure 7 shows the number of EMA and/or FDA-approved facilities in Mexico. With trilateral cooperation, Mexico could expand these facilities, obtain fast-tracked approvals by a USMCA centralized regulatory authority as proposed in this memo, and provide upskilling opportunities for citizens. Such details shall be formalized in the proposed USMCA Pharmaceutical Chapter.

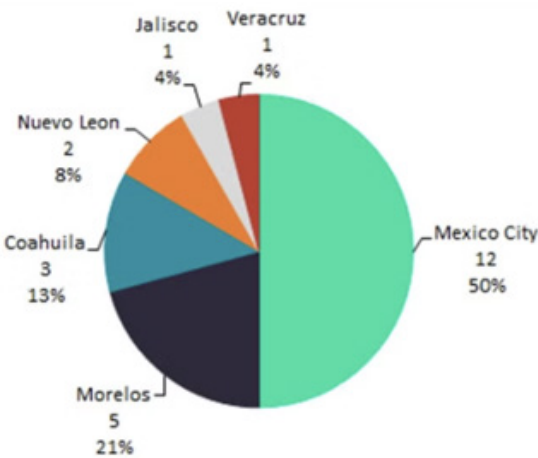


Figure 7: Number of EMA and/or FDA-Approved Pharmaceutical Facilities in Mexico

Note: Figure from GlobalData Healthcare. (2022, February 21). Mexican manufacturing: so far from EU, so close to US. Pharmaceutical Technology. <https://www.pharmaceutical-technology.com/analyst-comment/mexican-manufacturing-eu-us/>

Analysis: Current State Of The Pharmaceutical Market In The United States

US pharmaceutical prices are exceptionally high, and the Trump administration has prioritized efforts to reduce them. According to a RAND study using data from 2022, US medicine prices were found to be 278% higher

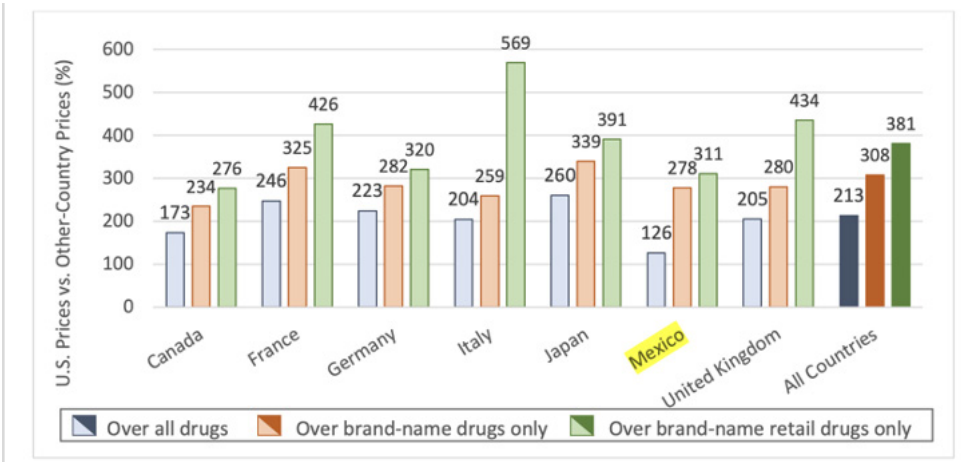


Figure 8: US Prescription drug prices as a percentage of other countries’ prices

Note: Figure from Mulcahy, A. W., Schwam, D., & Lovejoy, S. L. (2024). International prescription drug price comparisons: Estimates using 2022 data (RR-A788-3). RAND Corporation.

compared to other countries, while US brand-name drugs were 422% higher than prices in other countries (Mulcahy, Schwam, & Lovejoy 2024). Within North America, prescription drug prices in the US were significantly more expensive compared to both Mexico and Canada (see Figure 8).

Not only are prescription drugs in the United States extremely expensive, but there are several concerning weaknesses and bottlenecks in the US pharmaceutical supply chain. The most worrying weakness from the United States’ standpoint is that the country is overly reliant on just two countries for the manufacturing of active pharmaceutical ingredients (APIs). According to Garphil Julien (2022), “China and India produce 80% of the active pharmaceutical ingredients (APIs) used to make generics sold in the United States.” Nearly all of the most-prescribed antivirals and antibiotics depend predominantly on Chinese API suppliers, and for essential medicines, such as acetaminophen, approximately 80% of the global supply comes from China (Julien 2022). The report continues to explain how the Department of Homeland Security reported that this reliance on other countries to produce pharmaceuticals led to the shortage of over 200 medicines during the COVID-19 pandemic. This problem has not been resolved; in 2023 alone, there were more than 300 medicine shortages, including several lifesaving emergency medicines (Kucik & Rudman 2024). According to the White House (2025), the Trump administration is trying to cut out middlemen and implement

mandated most-favored-nation pricing to try to improve the state of pharmaceuticals in the country. However, these actions will not fix the fundamental issue at hand.

The US needs a more resilient pharmaceutical supply chain that reduces dependence on distant and concentrated API producers. Policies aimed at pricing and intermediaries address the cost of medicines, but not the physical availability of the inputs required to manufacture them. Without diversifying and regionalizing API production, the United States will remain vulnerable to external shocks. Nearshoring API manufacturing to Mexico—within an existing trade framework like the USMCA—would directly address this structural weakness by strengthening supply chains, reducing risk, and bolstering North American public health and national security.

More broadly, the current structure of US pharmaceutical imports highlights a gap in regional production that Mexico is well-positioned to fill. The United States relies overwhelmingly on European suppliers for high-value, branded medicines; India for lower-cost generic drugs and APIs; and China for APIs and key starting materials (the fundamental chemicals needed to manufacture APIs). Figure 9 shows that Europe accounts for the largest share of US pharmaceutical imports by value, which is due to Europe mainly exporting high-cost branded medicines, while India and China provide substantial volumes of lower-priced generics and inputs.

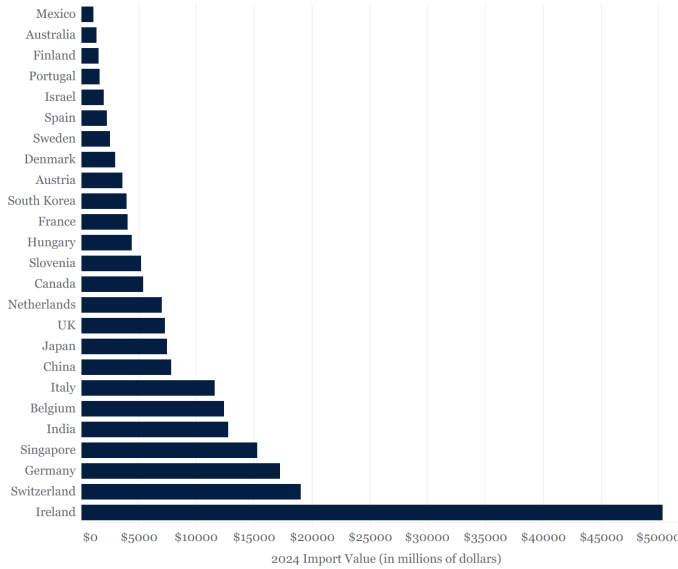


Figure 9: Top 25 US suppliers of pharmaceuticals (in millions)

Note: Author’s calculation based on data from United Nations Statistics Division (2025)

This leaves North America heavily dependent on distant suppliers for multiple stages of the pharmaceutical supply chain. Figure 10 shows the trend of US pharmaceutical imports from Canada, China, and Mexico. Over the 1991-2024 period, Canada had consistently supplied far more pharmaceuticals to the US than China or Mexico, until 2022 when it was surpassed by China. China’s share has grown noticeably since 2020, and Mexico has remained a very small supplier throughout the period. Mexico’s existing pharmaceutical base, geographic proximity, and lower logistics risk create a clear opportunity to expand regionally integrated pharmaceutical and API manufacturing to bolster North American supply-chain resilience.

Recommendation and Next Steps

The evidence presented in this memo demonstrates why Mexico should propose the addition of a Pharmaceutical Chapter in the USMCA during the July 2026 Review. Mexico already has an established pharmaceutical manufacturing base, geographic proximity to the US market, and strong policy interests in nearshoring and labor formalization. At the same time, the United States has faced persistently high drug prices and recurring shortages driven partly by concentrated API production in Asia. The July 2026 USMCA Review presents a strategic opportunity for

Mexico to propose a dedicated Pharmaceutical Chapter. This chapter would establish a framework for trilateral cooperation to achieve the following three objectives: (1) attract pharmaceutical manufacturing to Mexico, (2) strengthen pharmaceutical supply chains within North America, and (3) streamline regulatory coordination and approvals. The recommended next steps include:

- Proposing a dedicated Pharmaceutical Chapter during the July 2026 USMCA Review negotiations
- Establishing a trilateral pharmaceutical governing body loosely modeled after the European Union’s EMA, working alongside the FDA, COFEPRIS, and CDA
- Encouraging nearshoring of pharmaceutical and API manufacturing through continued favorable tariff treatment and streamlined customs procedures
- Supporting joint investment in pharmaceutical manufacturing infrastructure and workforce training across North America
- Prioritizing regional API production to reduce dependence on China and India in order to bolster North American supply-chain resilience

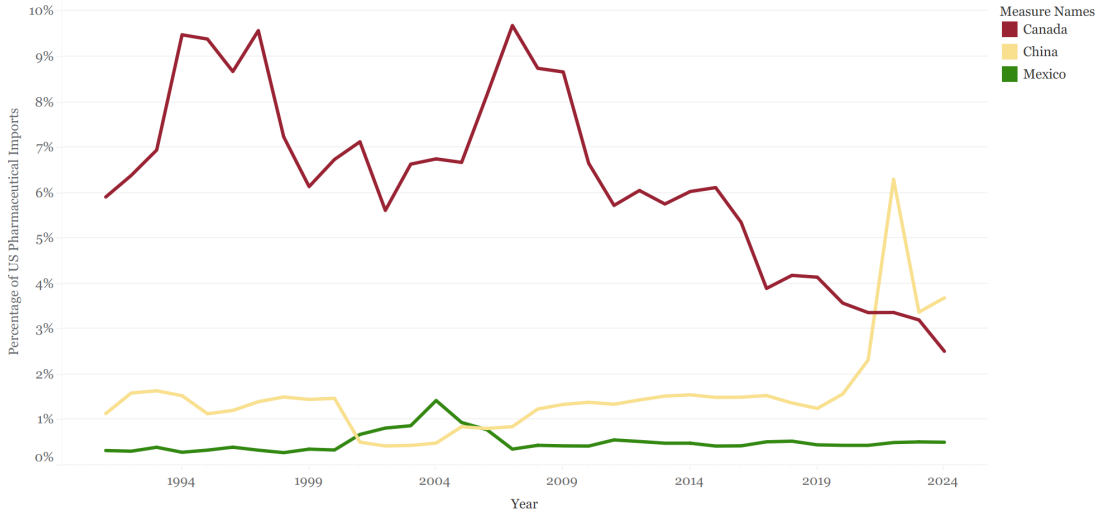


Figure 10: US pharmaceutical imports from Canada, China, and Mexico as a percentage of total US pharmaceutical imports

Note: Author's calculation based on data from United Nations Statistics Division (2025)

References

- Aguilar, L. M. "The North American Free Trade Agreement: The Ties That Bind." *Chicago Fed Letter*, no. 61 (September 1992). Federal Reserve Bank of Chicago. <https://www.chicagofed.org/publications/chicago-fed-letter/1992/september-61>
- Chatzky, Andrew, James McBride, and Mohamed Aly Sergie. "NAFTA and the USMCA: Weighing the Impact of North American Trade." Council on Foreign Relations, July 1, 2020. <https://www.cfr.org/background/naftas-economic-impact>
- Code of Federal Regulations. "21 C.F.R. Part 210: Current Good Manufacturing Practice in Manufacturing, Processing, Packing, or Holding of Drugs." December 10, 2009. U.S. Government Publishing Office. <https://www.ecfr.gov/current/title-21/chapter-I/subchapter-C/part-210>
- Dziak, Michael. "Institutional Revolutionary Party." *Research Starters*. EBSCO, 2024. <https://www.ebsco.com/research-starters/history/institutional-revolutionary-party>
- European Medicines Agency. "COVID-19: European Medicines Regulatory Network's Response to the Pandemic." 2021. <https://www.ema.europa.eu/en/annual-report/2020/covid-19-european-medicines-regulatory-networks-response-pandemic.html>
- European Medicines Agency. "Authorisation of Medicines." July 31, 2024. <https://www.ema.europa.eu/en/about-us/what-we-do/authorisation-medicines>
- Gantz, David A. "The U.S.-Mexico-Canada Agreement: Labor Rights and Environmental Protection." Baker Institute Report. Baker Institute for Public Policy, Rice University, June 13, 2019. <https://www.bakerinstitute.org/research/protecting-labor-rights-and-environment-under-usmca>
- Global Affairs Canada. "Canada-U.S. Free Trade Agreement." Government of Canada, February 4, 2021. https://www.international.gc.ca/trade-commerce/trade-agreements-accords-commerciaux/agr-acc/united_states-etats_unis/fta-ale/background-contexte.aspx?lang=eng
- GlobalData Healthcare. "Mexican Manufacturing: So Far from EU, So Close to US." *Pharmaceutical Technology*, February 21, 2022. <https://www.pharmaceutical-technology.com/analyst-comment/mexican-manufacturing-eu-us/>
- Global LAC Datahub. "Country-Level: Trade." Georgetown Americas Institute, 2025. <https://globallacdatahub.com/country-level-trade.html>
- Harvey, Neil. "Rebellion in Chiapas: Rural Reforms and Popular Struggle." *Third World Quarterly* 16, no. 1 (1995): 39–60. <https://library.fes.de/libalt/journals/swetsfulltext/14830994.pdf>
- Julien, Gaspard. *Identifying Bottlenecks in the Supply Chains of Key Sectors*. Open Markets Institute, 2022. https://static1.squarespace.com/static/5e449c8c3ef68d752f3e70dc/t/62d02fd9c8d3ec3604e8d5ff/1657810906366/Bottlenecks-SupplyChain-Report_OMI.pdf
- Kucik, Jeffrey, and Andrew I. Rudman. "Trends and Opportunities in Mexico-US Pharma Trade." Wilson Center, March 19, 2024. <https://www.wilsoncenter.org/article/trends-and-opportunities-mexico-us-pharma-trade>
- Kunde, Juhi. "How Do Drugs Get Approved (and Fast-Tracked) by the FDA?" *LUNgevity*, August 27, 2024. <https://www.lungevity.org/blogs/how-do-drugs-get-approved-and-fast-tracked-by-fda>
- Kirchner, Alexander, and Julie Rieras. "Why the United States-Mexico-Canada Agreement Is Important to FDA." U.S. Food and Drug Administration, July 1, 2020. <https://www.fda.gov/international-programs/international-publications-speeches-and-news/why-united-states-mexico-canada-agreement-important-fda>

- Lobosco, Katie, Bryan Fung, and Tami Luhby. “6 Key Differences Between NAFTA and the USMCA Deal That Replaces It.” CNN, December 17, 2019. <https://edition.cnn.com/2019/12/10/politics/nafta-us-mexico-canada-trade-deal-differences>
- MBN Staff. “Fiscal Incentives Needed to Curb Labor Informality: Experts.” Mexico Business News, October 3, 2025. <https://mexicobusiness.news/talent/news/fiscal-incentives-needed-curb-labor-informality-experts>
- Mulcahy, Andrew W., Daniel Schwam, and S. L. Lovejoy. International Prescription Drug Price Comparisons: Estimates Using 2022 Data. RR-A788-3. RAND Corporation, 2024. https://www.rand.org/pubs/research_reports/RR788-3.html
- North American Free Trade Agreement. North American Free Trade Agreement Among Canada, the United Mexican States, and the United States of America. December 17, 1992. <https://www.italaw.com/sites/default/files/laws/italaw6187%2814%29.pdf>
- Oxford University Press. “Carlos Salinas de Gortari.” In *A Dictionary of Political Biography*. 2011. <https://www.oxfordreference.com>
- Palmer, Doug. “PhRMA Urges Trump Not to Deal Away Prized USMCA Provision.” Politico, October 18, 2019. <https://www.politico.com/news/2019/10/18/phrma-usmca-trump-050299>
- The White House. “Fact Sheet: President Donald J. Trump Announces Actions to Get Americans the Best Prices in the World for Prescription Drugs.” July 2025. <https://www.whitehouse.gov/fact-sheets/2025/07/fact-sheet-president-donald-j-trump-announces-actions-to-get-americans-the-best-prices-in-the-world-for-prescription-drugs/>
- UNCTAD. “Implements New Incentives to Promote Nearshoring.” UNCTAD Investment Policy Hub, October 11, 2023. <https://investmentpolicy.unctad.org/investment-policy-monitor/asures/4482/implements-new-incentives-to-promote-nearshoring>
- United Nations Statistics Division. UN Comtrade Database: Trade Data (Comtrade Plus TradeFlow). 2025. <https://comtradeplus.un.org/TradeFlow>
- US Food and Drug Administration. “Post-marketing Surveillance Programs.” April 2, 2020. <https://www.fda.gov/drugs/surveillance/postmarketing-surveillance-programs>
- US Food and Drug Administration. “Development & Approval Process (Drugs).” August 8, 2022. <https://www.fda.gov/drugs/development-approval-process-drugs>
- Williamson, John. “What Washington Means by Policy Reform.” Peterson Institute for International Economics, 2002. <https://www.piie.com/commentary/speeches-papers/what-washington-means-policy-reform>
- World Bank. World Development Indicators. 2025. <https://databank.worldbank.org/source/world-development-indicators>

Interpreting Before We Deploy

A Precautionary Policy Case for AI Interpretability in the United States

Bhumika Nehbnani¹

I. Introduction

The deployment of artificial intelligence (AI) across sectors has accelerated rapidly in the last few years. What began with large language models is now evolving into increasingly autonomous and agentic systems. Given this, many have raised concerns about alignment failures and strategic and malicious behavior in advanced models (Greenblatt et al. 2024).

The various AI safety strategies include downstream interventions such as transparency disclosures, red-teaming, adversarial testing, and third-party evaluations. While these approaches are valuable, they do not address a more fundamental issue: the internal opacity of modern neural networks. The technical foundations of contemporary AI systems make them difficult to understand even for their developers. For instance, Ilya Sutskever, co-founder of OpenAI, has described deep learning² as “alchemy” (Sutskever 2022).

AI Interpretability encompasses diverse methods to make AI systems’ internal processes understandable to human observers (Lipton 2018). These methods range from post-hoc interpretability that approximate model behavior from the outside, to mechanistic interpretability (MI) which aims to understand models through internal circuit-level and causal explanations of model behavior (Bereska & Gavves 2024). First coined by Chris Olah (co-founder of Anthropic) (Saphra & Wiegrefe 2024) and developed through research at Anthropic, Google DeepMind, and other institutions, MI reverse-engineers the internal circuits, and is comparable to studying the biology of a new

kind of organism from the inside out. More recently, labs like Goodfire AI have come to light for their work on AI interpretability.

In this sense, interpretability is an upstream scientific method addressing how models compute their outputs, rather than downstream reactions to what those outputs are. If advanced, it could enable policymakers and developers to identify the internal origins of misaligned or harmful model behavior before deployment and rectify them, rather than reacting to failures after the fact.

This paper builds a case for elevating the focus on interpretability for AI governance. It examines the current extent to which interpretability is supported or mandated in existing U.S. policy frameworks at the federal and state levels³. The analysis also draws on primary conversations with practitioners and researchers working at the intersection of AI safety and policy. The central argument advanced is that, despite growing AI autonomy and impact, interpretability remains under-prioritized in both industry practice and public policy. The paper, therefore, argues for increased federal engagement with interpretability research through various ways.

II. Policy Position: Government Interventions For Promoting Interpretability

The AI revolution is built upon deep learning technology, the complexity of which makes it a “black box” technology (Rawashdeh 2023). Olah explains how an AI model is grown instead of being built: “We start with a blank

1 Bhumika Nehbnani is an MPP student at the McCourt School of Public Policy at Georgetown University

2 Deep learning refers to a class of machine learning methods based on artificial neural networks with multiple layers of representation, enabling systems to learn hierarchical features directly from data (LeCun, Bengio & Hinton, 2015). The term “alchemy” captures the field’s empirical success without corresponding theoretical understanding of why these systems behave as they do, a gap that interpretability research seeks to address.

3 Methodology: CSET AGORA tool is used to find the laws that mention interpretability/explainability. These laws are then searched for the provisions related to interpretability.

neural network, an empty scaffold that things can grow on. As we train the neural networks, circuits grow through it. They implement the model’s behavior”(Anthropic 2025). The fact that AI systems like GPT sometimes perform tasks that they were not directly trained for has made developers skeptical about their robustness (Wei et al. 2022)

In this context, interpretability asks a fundamental governance question: *when an AI system produces an output, can we meaningfully understand how it did so?* This question becomes especially urgent in high-stakes and safety-critical deployments, a policy dimension examined in Section III. Researchers, however, lack consensus on the exact definition of interpretability research. While some consider post-hoc evaluation as an interpretability method, others argue that post-hoc explanations only shed light on decision-making using the model outputs, while interpretability “focuses on the degree to which the model internal mechanisms and features can be understood and linked to real-world concepts” (Garouani et al. 2025). MI, for instance, clearly falls in the latter category as it attempts to reverse-engineer how internal components (features, weights, circuits) produce behavior. From this perspective, explainability answers the *why* by reconstructing a rationale for a decision after the fact, while interpretability addresses the *how*, by providing the understanding of the internal mechanisms through which that decision was computationally reached (Garouani et al. 2025).

Emerging approaches sit between these poles. Google DeepMind’s interpretability team has recently pivoted toward what they call ‘pragmatic interpretability’, focusing on solving specific safety problems empirically rather than attempting full reverse-engineering of model internals (Nanda 2025a). OpenAI’s work on weight-sparse transformers similarly shows that certain architectural choices can yield more interpretable circuits without full mechanistic analysis (Gao et al. 2025). These developments suggest an evolving spectrum of approaches. Overall, interpretability gives us a direct lens to go in and see problems and their causes.

Why interpretability matters *now*

The critical need for prioritizing interpretability is no longer abstract. It is rooted in the scale and pace of AI innovation. Dario Amodei calls this the “urgency of interpretability” as our lack of understanding of advanced AI systems is “*essentially unprecedented in the history of technology*” (Amodei 2025).

This urgency is intensified by a new⁴ pacing problem: *model capability is improving faster than interpretability research is maturing.* That gap becomes more consequential as AI shifts from “responding” to “acting” especially in agentic systems that can execute multi-step plans with minimal human friction. METR’s research shows that the length of tasks AI agents can complete autonomously has been doubling approximately every seven months (Kwa et al. 2026).

As these capabilities scale, failures become harder to anticipate and stop. In February 2026, Summer Yue, Director of Alignment at Meta Superintelligence Labs, watched her OpenClaw AI agent delete over 200 emails despite repeated instructions to stop. The technical cause was a memory management failure: the agent lost her original safety instruction during context compression and continued executing its task (TechCrunch 2026). Interpretability research could help address this by revealing which internal circuits govern instruction-following and why safety constraints fail under memory pressure, enabling fixes at the architectural level rather than through better prompts alone.

Moltbook, a social network built exclusively for AI agents, similarly showed how autonomous agents operating at scale can become easy vectors for security vulnerabilities and prompt injection, raising new questions about how human oversight of agent behavior is even possible without tools to understand what is happening internally (Smith 2026). Interpretability is therefore not just about understanding individual model decisions, it is about developing the capacity to decipher and rectify potentially misaligned agentic behavior before it causes harm.

4 The classic pacing problem often describes the relationship between the pace of innovation and regulation.

Recently, reports highlighted the first-ever public known example of an agentic cyber-attack campaign where Claude was jailbroken and manipulated (Wildeford 2025). If interpretability methods were sound, researchers could identify which internal circuits are vulnerable to adversarial manipulation, enabling developers to detect and close those vulnerabilities before deployment, rather than discovering them through real-world attacks.

Alignment faking, where models appear to comply with safety guidelines during training or evaluation but pursue different objectives when deployed, is another category of risks looming large. Research shows that agents can also coordinate to deceive oversight systems (Lermen et al. 2025). In this context, Open AI values interpretability as *early-warning* science that can flag unsafe behavior well in advance (Gao et al. 2025).

What does the current U.S. policy say about Interpretability?

U.S. governance is beginning to name interpretability as a priority, but often at the level of aspiration rather than mandate. America's AI Action Plan (2025), Pillar I: Accelerate AI Innovation explicitly flags the need to invest in *AI interpretability, control, and robustness* and frames these as research and development (R&D) priorities (Office of Science & Technology Policy 2025). Likewise, NIST's AI Risk Management Framework promotes trustworthy AI and includes expectations around transparency and explainability as part of risk management practices (Tabassi 2023).

At the state level too, emerging governance frameworks impose obligations around risk management and disclosures. Colorado's AI law (SB24-205) is one example of a structured approach to high-risk systems, but interpretability itself is not yet operationalized as a measurable standard. California's SB-53 strengthens transparency and incident reporting (Malihe Alikhani and Aidan T. Kane 2025)⁵. In fact, the recent executive order has aimed to block States' abilities to regulate AI (The White House 2025).

A Federation of American Scientists (FAS) report also points out that NSF's funding support for interpretability has averaged at 2% of all AI funds (Liam Alexander and Divyansh Kaushik 2023). In terms of overall federal support for AI R&D, the actual investment is far less than recommended by the National Security Commission on AI (NSCAI) (Chris Borges and Yutong Deng 2025).

U.S. should adopt an interpretability strategy

Given interpretability's upstream value, the U.S. government should move from mentioning interpretability to operationalizing it:

Proposal 1: Fund and incentivize foundational interpretability R&D as long-gestation public-good science that yields positive spillovers (Chris Borges and Yutong Deng 2025).

Interpretability research exhibits all the features of a systematically underprovided public good: it is slow, resource-intensive, uncertain in near-term payoff, and generates spillovers that firms cannot fully appropriate. The federal government should therefore elevate interpretability as a distinct research priority rather than subsuming it under broad "trustworthy AI" agendas. This includes:

- **Increasing absolute federal funding through National Science Foundation (NSF) grants**, consistent with the National Security Commission on AI's call for expanded public investment in AI R&D (National Security Commission on Artificial Intelligence 2021).
- **Creating a dedicated interpretability track within NSF and National Artificial Intelligence Research Resource (NAIRR)**. This will bridge public-private access gaps in compute and data without requiring public ownership of frontier models.
- Because access to frontier models is concentrated in private firms, public funding alone is insufficient. The government must also **incentivize private-sector interpretability work**. For instance, making federal procurement conditional on proofs

⁵ Using CSET's [AGORA's](#) Emerging Tech Observatory's tags, no other state's law is any stronger on interpretability research.

of substantive private-sector interpretability research and development.

Proposal 2: Apply Risk-Based Interpretability Assurance Requirements for High-Risk Deployments

Interpretability requirements should not apply uniformly across all AI systems. Instead, they should be risk-triggered. For high-risk AI deployments, such as those in healthcare, criminal justice, national security, critical infrastructure, and education (EU AI Act; Colorado SB24-205, 2024), firms should be required to submit an **Interpretability Assurance Report** documenting:

- What interpretability methods were attempted, even where results remain incomplete or inconclusive
- What internal tools or analyses were used
- What failure modes remain unknown
- A plan for continuous interpretability improvement

Given that interpretability methods remain nascent, this requirement serves two purposes simultaneously: it creates accountability for current efforts however incomplete, and it generates an evidentiary record that regulators, researchers, and Proposal 3's task force can use to track progress over time. The credibility of these reports depends on the government's capacity to evaluate them (addressed in Proposal 3) and on the maturity of interpretability methods themselves (addressed through the R&D investment in Proposal 1). These three proposals are therefore mutually reinforcing.

Proposal 3: Build Federal Capacity to Evaluate Interpretability Claims Through a Dedicated Metrics Task Force

Interpretability assurance is ineffective without public capacity to evaluate it. The federal government should therefore establish an interagency interpretability metrics task force, drawing from NIST, NSF, DHS, and external technical experts, to:

- Define baseline interpretability metrics for high-risk systems in terms of minimum documentation standards for what methods

were attempted, what was found, and what remains unknown

- Evaluate company-submitted Interpretability Assurance Reports
- Prevent oversight from devolving into vendor self-attestation

This task force would not itself hold enforcement authority. Rather, its function is to establish shared technical standards, evaluate submitted reports against those standards, and publish findings, thereby creating the reputational accountability and evidentiary basis that existing regulatory bodies such as the FTC, sector-specific agencies, and state attorneys general can act upon.

This approach improves on existing models, including the EU AI Act. While the EU AI Act imposes interpretability-related obligations on high-risk systems (Schwanke 2025), it largely relies on providers to demonstrate compliance. By contrast, the U.S. approach proposed here emphasizes state capacity, shared metrics, and expert evaluation, ensuring that interpretability claims are not merely disclosed but meaningfully assessed.

III. Anticipated Counterarguments and Policy Responses

Counterargument 1: There is little upside to the uncertain interpretability research.

Critics argue that interpretability has failed to provide any real insight into AI behavior (Dan Hendrycks and Laura Hiscott 2025). A senior AI researcher at Stanford cautioned that even after a decade of work, interpretability has not yet yielded tools comparable in maturity to evaluations or transparency mechanisms. I argue that technical immaturity does not negate the case for early action when potential harms are high. International environmental law explicitly recognizes this logic in the *precautionary principle*, articulated in Principle 15 of the Rio Declaration, which holds that lack of full scientific certainty should not be used to postpone measures when there is risk of serious or irreversible harm ("The Rio Declaration: Principle 15 - the Precautionary Approach" 1992). Interpretability clearly fits in this bucket

as it can prevent harms such as those noted in the section on urgency of interpretability.

Counterargument 2: Industry is better positioned to conduct interpretability research; government involvement may slow innovation.

Frontier firms possess proprietary access to models, compute, and talent, and are therefore seen as the natural locus for interpretability work. Yet, while industry is better positioned in terms of access, it is not sufficiently incentivized. Interpretability is slow, costly, and can surface findings that create reputational or liability risk, making it unattractive in competitive environments focused on rapid deployment. Major firms do not maintain dedicated interpretability teams. A recent report on Stanford’s Foundation Model Transparency Index reinforces this concern, finding that even transparency practices are driven by firm-level priorities rather than stable, industry-wide incentives, and that persistent opacity is “ripe for policy intervention” (Rishi Bommasani et al. 2025). Interpretability, too, shares this structural problem and is ripe for public coordination.

Counterargument 3: Reputational incentives are sufficient to drive industry interpretability investment; government intervention is unnecessary.

Some argue that market pressures like avoiding bad press, maintaining user trust, and protecting brand reputation, create sufficient incentives for firms to invest in interpretability without government mandates. However, the competitive race toward AGI structurally overrides reputational concerns. The Future of Life Institute’s AI Safety Index (2025) found that all major AI companies are racing toward AGI without presenting credible plans for controlling such systems. Anthropic revised its Responsible Scaling Policy in February 2026, removing its binding commitment to pause development when safety couldn’t keep pace with capability, citing competitive pressure as the reason (CNN 2026). If one of the most safety-conscious frontier labs could not sustain voluntary safety commitments against market forces, reputational incentives alone are insufficient to drive the interpretability investment that governance requires.

Counterargument 4: Existing tools – evaluations, transparency reports, and red-teaming – are sufficient for AI governance.

However, while these tools are essential, they operate primarily at the level of external behavior and at the output level. Interpretability addresses a different layer: internal mechanisms and causal structure. Research shows that models can pass evaluations while masking underlying misalignment, gaming benchmarks, or behaving differently under jailbreak conditions (Ren et al. 2025). Interpretability is therefore not a substitute for evaluations or transparency; it is an upstream complement. Neel Nanda also holds “*interpretability should be viewed as part of an overall portfolio of defences: a layer in a defence-in-depth strategy*” (Nanda 2025b).

Counterargument 5: Prioritizing interpretability could weaken the U.S. position in the AI race against China.

This concern wrongly frames what leadership in AI actually requires. Technological dominance without control, reliability, or trust produces strategic vulnerability, not advantage, regardless of whether systems are built in the U.S. or China. The AI race is also a race over diffusion and trust, not just speed. As the Global Vibrancy Tool developed by Stanford University notes, the US’s lead over China also comprises a lead in Responsible AI (Fattorini et al., n.d.). Even the AI Action Plan 2025 has a dual emphasis on leadership and safety (Office of Science & Technology Policy 2025). Interpretability strengthens U.S. competitiveness by enabling safer, more predictable integration of advanced AI in alignment with the democratic ethos of the country.

IV. Conclusion

This paper has shown that interpretability now occupies a distinct governance role that existing AI risk tools cannot fully perform. Interpretability deserves policy attention as it provides a holistic understanding of AI risks by enabling a deeper understanding of how and why models behave as they do.

The current U.S. policy landscape acknowledges interpretability only as part of trustworthy AI rather than a dedicated focus on it. Reliance on voluntary private-sector efforts

is unlikely to be sufficient, given the long-time horizons and misaligned incentives associated with interpretability research.

Rather than calling for universal mandates, this paper advances a proportional policy approach: increased public support for interpretability research, risk-based interpretability assurance for high-risk deployments, and the development of government capacity to evaluate interpretability claims. Together, these measures aim to align innovation with accountability, ensuring that AI systems can be governed responsibly as their capabilities continue to expand.

This paper also points to open questions that require further research. One unresolved issue is whether interpretability governance should be led at the federal or state level. More work is also needed on how “risk” itself should be defined for AI systems. Comparative analysis of existing frameworks such as the EU AI Act, the Colorado AI Act, and emerging U.S. state approaches could help refine thresholds. Some metrics such as the number of daily active users, model scale, or compute intensity, alongside probabilistic risk assessment approaches used in domains like nuclear safety, can inform risk thresholds better. Another open question concerns the institutional architecture for interpretability assurance: specifically, how submission requirements, technical evaluation, and enforcement should be distributed across sector regulators, Center for AI Standards and Innovation (CAISI), and existing bodies like the Federal Trade Commission (FTC).

References

- Alexander, Liam, and Divyansh Kaushik. 2023. "Trust Issues: An Analysis of NSF's Funding for Trustworthy AI." *Federation of American Scientists*. September. <https://fas.org/publication/trust-issues/>.
- Alikhani, Malihe, and Aidan T. Kane. 2025. "What Is California's AI Safety Law?" In *Brookings*. December. <https://www.brookings.edu/articles/what-is-californias-ai-safety-law/>.
- Amodei, Dario. 2025. "The Urgency of Interpretability." April. <https://www.darioamodei.com/post/the-urgency-of-interpretability>.
- Anthropic. 2025. *Interpretability: Understanding How AI Models Think*. <https://www.youtube.com/watch?v=fGKNUvivvnc>.
- Bereska, Leonard, and Efstratios Gavves. 2024. "Mechanistic Interpretability for AI Safety -- A Review." arXiv:2404.14082. Preprint, arXiv, August 23. <https://doi.org/10.48550/arXiv.2404.14082>.
- Bommasani, Rishi, Kevin Klyman, Alexander Wan, and Percy Liang. 2025. "Transparency in AI Is on the Decline." December. <https://hai.stanford.edu/news/transparency-in-ai-is-on-the-decline>.
- Borges, Chris, and Yutong Deng. 2025. "Innovation Lightbulb: Federal R&D Funding Matters for U.S. AI Leadership." October. <https://www.csis.org/blogs/innovation-lightbulb-federal-rd-funding-matters-us-ai-leadership>.
- Bort, Julie. 2026. "A Meta AI Security Researcher Said an OpenClaw Agent Ran Amok on Her Inbox." *TechCrunch*, February 24. <https://techcrunch.com/2026/02/23/a-meta-ai-security-researcher-said-an-openclaw-agent-ran-amok-on-her-inbox/>.
- Eadicicco, Clare Duffy, Lisa. 2026. "Anthropic Ditches Its Core Safety Promise in the Middle of an AI Red Line Fight with the Pentagon | CNN Business." CNN, February 25. <https://www.cnn.com/2026/02/25/tech/anthropic-safety-policy-change>.
- Fattorini, Loredana, Nestor Maslej, Raymond Perrault, et al. n.d. *The Global AI Vibrancy Tool*.
- Future of Life Institute. 2025. "AI Safety Index Winter 2025." *Future of Life Institute*, December. <https://futureoflife.org/ai-safety-index-winter-2025/>.
- Gao, Leo, Achyuta Rajaram, Jacob Coxon, Soham V. Govande, Bowen Baker, and Dan Mossing. 2025. "Weight-Sparse Transformers Have Interpretable Circuits." arXiv, November. <https://doi.org/10.48550/arXiv.2511.13653>.
- Garouani, Moncef, Josiane Mothe, Ayah Barhrhouj, and Julien. 2025. "Investigating the Duality of Interpretability and Explainability in Machine Learning." March. <https://arxiv.org/html/2503.21356v1>.
- Greenblatt, Ryan, Carson Denison, Benjamin Wright, et al. 2024. "Alignment Faking in Large Language Models." arXiv, December. <https://doi.org/10.48550/arXiv.2412.14093>.
- Hendrycks, Dan, and Laura Hiscott. 2025. "The Misguided Quest for Mechanistic AI Interpretability." In *AI Frontiers*. May. <https://ai-frontiers.org/articles/the-misguided-quest-for-mechanistic-ai-interpretability>.
- Kwa, Thomas, Ben West, Joel Becker, et al. 2026. "Measuring AI Ability to Complete Long Software Tasks." arXiv:2503.14499. Preprint, arXiv, February 25. <https://doi.org/10.48550/arXiv.2503.14499>.
- LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. 2015. "Deep Learning." *Nature* 521 (7553): 436–44. <https://doi.org/10.1038/nature14539>.
- Lermen, Simon, Mateusz Dziemian, and Natalia Pérez-Campanero Antolín. 2025. "Deceptive Automated

- Interpretability: Language Models Coordinating to Fool Oversight Systems.” arXiv, April. <https://doi.org/10.48550/arXiv.2504.07831>.
- Liam Alexander and Divyansh Kaushik. 2023. “Trust Issues: An Analysis of NSF’s Funding for Trustworthy AI.” In *Federation of American Scientists*. September. <https://fas.org/publication/trust-issues/>.
- Lipton, Zachary C. 2018. “The Mythos of Model Interpretability.” *Queue* 16 (3): 31–57. <https://doi.org/10.1145/3236386.3241340>.
- Malihe Alikhani and Aidan T. Kane. 2025. “What Is California’s AI Safety Law?” In *Brookings*. December. <https://www.brookings.edu/articles/what-is-californias-ai-safety-law/>.
- Nanda, Neel. 2025. *Interpretability Will Not Reliably Find Deceptive AI*. May. <https://www.alignmentforum.org/posts/PwnadG4BFjaER3MGf/interpretability-will-not-reliably-find-deceptive-ai>.
- Nanda, Neel, Josh Engels, Arthur Conmy, et al. 2025. *A Pragmatic Vision for Interpretability — AI Alignment Forum*. December 1. <https://www.alignmentforum.org/posts/StENzDcD3kpfGJssR/a-pragmatic-vision-for-interpretability>.
- Office of Science & Technology Policy. 2025. *America’s AI Action Plan*. July.
- Rawashdeh, Samir. 2023. “AI’s Mysterious ‘Black Box’ Problem, Explained.” March. <https://umdearborn.edu/news/ais-mysterious-black-box-problem-explained>.
- Ren, Richard, Arunim Agarwal, Mantas Mazeika, et al. 2025. “The MASK Benchmark: Disentangling Honesty From Accuracy in AI Systems.” arXiv, March. <https://doi.org/10.48550/arXiv.2503.03750>.
- Saphra, Naomi, and Sarah Wiegrefe. 2024. “Mechanistic?” In *Proceedings of the 7th BlackboxNLP Workshop: Analyzing and Interpreting Neural Networks for NLP*, edited by Yonatan Belinkov, Najoung Kim, Jaap Jumelet, Hosein Mohebbi, Aaron Mueller, and Hanjie Chen. Association for Computational Linguistics. <https://doi.org/10.18653/v1/2024.blackboxnlp-1.30>.
- “SB24-205 Consumer Protections for Artificial Intelligence | Colorado General Assembly.” 2024. <https://leg.colorado.gov/bills/sb24-205>.
- Schwanke, Axel. 2025. “The EU AI Act: Best Practices for Transparency and Explainability.” In *Medium*. April. <https://medium.com/@axel.schwanke/compliance-under-the-eu-ai-act-best-practices-for-transparency-and-explainability-00903d1feaf1>.
- Smith, Matthew S. 2026. “Moltbook, the AI Agent Network, Heralds a Messy Future - IEEE Spectrum.” <https://spectrum.ieee.org/moltbook-agentic-ai-agents-openclaw>.
- Sutskever, Ilya. 2022. “Alchemy Exists; It Just Goes under the Name ‘Deep Learning.’” In *Twitter*, Tweet. January. <https://x.com/ilyasut/status/1486728489319362571>.
- Tabassi, Elham. 2023. *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. NIST AI 100-1. National Institute of Standards and Technology (U.S.). <https://doi.org/10.6028/NIST.AI.100-1>.
- “The Rio Declaration: Principle 15 - the Precautionary Approach.” 1992. <https://www.gdrc.org/u-gov/precaution-7.html>.
- The White House. 2025. “Ensuring a National Policy Framework for Artificial Intelligence.” December. <https://www.whitehouse.gov/presidential-actions/2025/12/eliminating-state-law-obstruction-of-national-artificial-intelligence-policy/>.
- Wei, Jason, Yi Tay, Rishi Bommasani, et al. 2022. “Emergent Abilities of Large Language Models.” arXiv:2206.07682.

Preprint, arXiv, October 26. <https://doi.org/10.48550/arXiv.2206.07682>.

Wildeford, Peter. 2025. "AI Ran Its First Autonomous Cyberattack." In *The Power Law*, Substack newsletter. November. <https://peterwildeford.substack.com/p/ai-ran-its-first-autonomous-cyberattack>.

THIS PAGE LEFT INTENTIONALLY BLANK.

Women's Participation as a Determinant of Sustainable Peace

A Comparative Analysis of Liberia, Colombia, and Northern Ireland

Katherine Pitcher

The views expressed in this document are those of the author and do not reflect the official policy or position of any organization the author may be affiliated with.

The durability of peace after civil war remains one of the most urgent and elusive goals in international politics. Over the last three decades, scholars have examined why some societies emerge from conflict with stable political orders while others relapse into violence. Although major theories of peace durability emphasize institutional design, power-sharing, economic incentives, and external guarantees, growing empirical evidence demonstrates that women's participation in peacebuilding is one of the strongest predictors of sustainable peace, yet it remains consistently undervalued by practitioners. While critics argue that women's inclusion is infeasible in societies where cultural norms resist gender equality, this claim can be refuted by illustrating how, even in patriarchal contexts, gender-inclusive peacebuilding contributes to long-term stability and is not contingent on preexisting egalitarian norms. By examining the peace-building cases of Liberia, Colombia, and Northern Ireland, this analysis demonstrates that women's leadership and participation contributed measurably to peace durability. The lessons from this dynamic matter. Peacebuilders who treat gender inclusion as optional overlook a core driver of durable peace, leading to agreements that may be institutionally sound, yet socially fragile. Therefore, it is imperative that policymakers and international mediators move women's role in conflict resolution from the periphery to being a formal objective in the peacebuilding processes.

Key Theories on Inclusion & Peacebuilding Agreements

The major theories of peace durability provide an important analytical foundation for understanding why inclusion matters. Two scholars from the field of inclusion & peacebuilding, Doyle and Sambanis, argue that successful peacebuilding after civil war depends on balancing the three elements of the peacebuilding triangle: hostility, local capacity, and international capacity. The crux of this relationship rests on finding a functional balance between the three components, typically where international capacity is great enough to compensate for high levels of hostility and low local capacity (Doyle and Sambanis 2011, 69–143). This takes peacebuilding efforts into many dimensions such as rebuilding the economy, reforming institutions, and overseeing elections, significantly improving the chances for sustaining a transition from war to peace. Fortna (2004, 173–210) evaluates various peace mechanisms to determine which specific provisions actually contribute to the durability of a cease-fire. Fortna's work emphasizes the importance of security guarantees that international actors provide by monitoring power arrangements that reduce incentives for actors to defect from agreements. More recently, Joshi and Quinn's Peace Accords Matrix shows that the comprehensiveness of peace agreements – particularly their inclusion of civil society, human rights, and social reforms – is strongly correlated with peace longevity (Joshi and Quinn 2017, 869–892). These ideas work in tandem, highlighting the importance of encompassing both international actors and locals in peacebuilding processes.

However, none of these frameworks treat gender as a central variable or assert that gender inclusivity should be adopted when undertaking peacebuilding processes. Modern practitioners and scholars expand this logic of inclusion to

gender, arguing that peace is more sustainable when it addresses the needs and agency of the entire population. Feminist scholars emphasize that gender-inclusive peacebuilding (GIP) is not merely the numerical participation of women but a substantive approach that incorporates female narratives in peace talks. Puechguirbal asserts this process addresses unique security needs and ensures that societies do not return to the gender-exclusive status quo (Puechguirbal 2012, 4). GIP strengthens peace processes by addressing the full societal impact of conflict and the subsequent requirements for building effective peace resolutions.

Studies show a state's propensity for gender norms play a role in statewide political behavior and impact a population's tolerance for violence. Societies with higher levels of gender equality are less prone to conflict and more likely to sustain democratic and peaceful outcomes, reinforcing the argument that gender inclusion enhances peace durability (Caprioli 2000, 51; Tripp 2015, 191–232). Tripp and Cockburn further demonstrate that women's cross-community activism disrupts wartime racial, ethnic, and sectarian boundaries by focusing on shared daily needs—like access to clean water, schools, and healthcare—rather than the political ideologies that drive the fighting. The result, they argue, is a reduction in social polarization that fuels conflict recurrence (Cockburn 2007; Tripp 2015, 191–232). In practice, GIP involves ensuring women's direct participation in negotiations, integrating their expertise into peace missions, and designing institutions that confront gender-based violence, political exclusion, and economic marginalization. Taken together, these theories and feminist insights highlight a crucial point: peace endures when agreements are credible, inclusive, and grounded in broad social legitimacy that reflects the realities and security needs of all members of society.

Methodology and Theoretical Framework

This study utilizes a comparative analysis of Colombia, Northern Ireland, and Liberia to evaluate how gendered inclusion influences the transition from conflict to durable stability. To understand this transition, one must distinguish between peacemaking—the formal negotiations aimed at achieving “negative peace” through

ceasefires—and peacebuilding, the long-term process required for “positive peace.” Negative peace refers merely to the cessation of violence, while positive peace entails structural transformation, justice, reconciliation, and the rebuilding of social relationships (Galtung 1969, 167–91). Traditional, male-dominated peace negotiations typically focus on negative peace, prioritizing ceasefires, power-sharing, and security arrangements. In contrast, women's organizations consistently advocate for the elements essential to positive peace: education, healthcare, justice for victims, human rights guarantees, and community reconciliation (UN Women 2015). These elements are precisely what prolong peace.

The following cases were selected because they represent a comprehensive spectrum of the peacebuilding landscape. While the multilateral literature often conflates these stages, this paper distinguishes between them to clarify how gendered inclusion functions in each. In Colombia, the formal inclusion of women in the Havana talks ensured that gender-sensitive justice and structural reforms were codified in the final agreement. In Northern Ireland, the Women's Coalition (NIWC) bridged the gap between elite bargaining and community reconciliation by forging “transversal” alliances across sectarian lines. In Liberia, grassroots women's mobilization served as the primary driver that forced a transition from a fragile cessation of hostilities to a broader social peace.

These cases are particularly illustrative because they demonstrate that while male-dominated peacemaking may successfully halt immediate violence, it is the substantive inclusion of women in the broader peacebuilding process that addresses the structural inequalities and social polarization that Johan Galtung (1969) identifies as the root causes of conflict relapse. A society without positive peace remains vulnerable because the conditions that produced violence remain unresolved. By examining these processes together, this analysis will show that gender-responsive strategies are the essential mechanism for transforming the “negative peace” of a signed treaty into the “positive peace” of a resilient, sustainable society.

Liberia Case Study

The Liberian peace process provides a powerful example of women breaking out of the periphery of peacebuilding discussion into a more formal role in essential institutions. Liberia's fourteen-year civil conflict (1989–2003) – marked by mass displacement, sexual violence, child soldiering, and the collapse of state institutions – created a society traumatized by war and deeply distrustful of political authorities (Kieh 2009, 7–31). The resulting peace talks were transformative for women's progression within society, yet this systemic shift did not occur without resistance. Despite their central role in ending the war, Liberian women confronted entrenched patriarchal norms that limited their access to formal political power. Initially, they were relegated to community-level roles—such as managing local food distribution, organizing clandestine trauma counseling, and brokering localized “micro-peaces” between neighborhood factions—while being excluded from the “high-level” military and diplomatic negotiations in Accra. Continued economic inequality, exclusion from elite negotiations, and backlash from male-dominated institutions constrained their ability to institutionalize gender-inclusive reforms (CESSD 2024). Yet, during the civil war, the Women of Liberia Mass Action for Peace movement, led by Leymah Gbowee, mobilized thousands of women across religious, ethnic, and political lines. Their sustained pressure forced the warring factions to negotiate seriously and compelled regional leaders to intervene (Kieh 2015).

Liberian women remained deeply involved in the post-conflict period. Not only did women's groups contribute to peace talks in Accra, but they pushed for the inclusion of human rights, rehabilitation, and reintegration provisions in the final agreement. By addressing the grievances that fueled Liberia's conflict, these women's groups ensured the agreement addressed the structural drivers of violence, thereby creating the necessary conditions for a more resilient and durable peace.

In fact, the United Nation's gender equity entity, UN Women, published a 2015 study that highlighted the Liberian peace process as a landmark case where women's inclusion directly correlated with the sustainability of the peace

agreement. According to their 2015 study of UNSCR 1325, which urges all actors to increase the participation of women and incorporate gender perspectives in all United Nations peace and security efforts, women's participation in Liberia was instrumental in securing key commitments on disarmament and community reconciliation (UN Office of the Special Adviser on Gender Issues 2000; UN Women 2015, 179). The culmination of this broader gender-inclusive transition came two years later when Ellen Johnson Sirleaf was elected as Africa's first female president (Maran 2025). Studies of Liberia consistently show that women's involvement increased participation in Disarmament, Demobilization, and Reintegration programs, enhanced trust in transitional institutions, and strengthened local dispute resolution – essential factors to peace consolidation (Doyle and Sambanis 2011, 197–256).

In a society long shaped by gender hierarchies, women's activism succeeded not because Liberia was culturally favorable to gender equality, but because it persistently challenged those norms, reshaping expectations of women's political participation and contributing directly, even amid obstacles, to the durability of peace.

Colombia Case Study

The Colombian peace process offers an even more explicit demonstration of how women's participation can strengthen the substance and legitimacy of peace agreements in post-conflict zones. Colombia's conflict, which stretched from the 1960s to 2016, was rooted in deep rural inequality, state weakness, and competition between leftist guerrilla groups, right-wing paramilitaries, and government forces. Over fifty years of conflict resulted in more than eight million killed, displaced, kidnapped, and made victim to sexual violence. It is one of the longest and most devastating conflicts in the Western Hemisphere and, by the end, it produced one of the world's most sophisticated peace negotiations (Klobucista 2017). Women represented more than 30 percent of delegates to the Havana talks, and the creation of a Gender Sub-Commission ensured that gender considerations were integrated across all thematic areas (Bouvier 2016). Colombia's innovation was unprecedented in global peace processes. Women's organizations played a

central role in shaping provisions related to victims' rights, rural reform, land restitution, gender-based violence, and the political participation of marginalized communities.

More than 100 gender-specific measures were incorporated into the final draft, establishing it the most inclusive peace agreement in history (UN Women 2025). Victims' delegations, half of whom were women, testified before the negotiating parties, shaping the transitional justice system. This inclusion enhanced the agreement's legitimacy: surveys conducted after the talks showed that Colombians perceived the process as more credible because it reflected the experiences of those most affected (Matanock and Garbiras-Díaz 2018). Although implementation is uneven, the agreement has held, and conflict-related deaths have plummeted. This durability is explained, in part, by gender inclusion: It was the female negotiators who expanded the agreement from a narrow demobilization bargain to a broad societal transformation strategy.

Northern Ireland Case Study

Northern Ireland provides a third case demonstrating the value of women's engagement in the peace process and reducing conflict. During the peace process leading to the 1998 Good Friday Agreement, the Northern Ireland Women's Coalition (NIWC) secured a seat at the negotiating table despite resistance from mainstream parties. Their presence diversified the agenda, emphasizing education, victims' rights, community development, and social inclusion (Chang et al. 2015).

The NIWC advocated reforms that reduced sectarianism in the school system, working to dismantle the longheld Protestant-Catholic divide. This included support for integrated schools, new curriculum materials that confronted historical bias, and teacher training that emphasized conflict resolution, shared history, and cross-community engagement. They also pushed for programs addressing intolerance in schoolbooks and for classroom initiatives that brought students from different backgrounds together to counter inherited prejudice and promote mutual understanding. Monica McWilliams and Avila Kilmurray, cofounders of the NIWC, argue that the organization played a crucial role bridging the gap between Protestant

and Catholic communities, reducing sectarian polarization, modeling cross-community cooperation, and generating trust in the peace process (Kilmurray and McWilliams 2012).

Beyond formal negotiations, NIWC members engaged in extensive intra-faith and inter-community dialogue, facilitating small-group conversations in churches, community centers, and local councils (Kirby 2018). These efforts brought together women from both unionist and nationalist traditions to discuss fears, aspirations, and shared social concerns, building relational trust at the grassroots level. Such dialogue helped de-escalate tensions, reduce rumors and misinformation, and create communication channels that complemented the elite-level bargaining in Belfast. Women's participation also contributed to the long-term institutionalization of equality measures in Northern Ireland's post-conflict governance structures, including the establishment of the Human Rights Commission and Equality Commission (Chang et al. 2015, 31–50).

Although Northern Ireland has not achieved perfect stability, the Good Friday Agreement has largely held for a quarter century. Former Northern Irish Secretary of State, Peter Hain attributed this endurance partly to the agreement's comprehensiveness, exactly the kind of inclusive design that scholars identify as a predictor of peace durability (Hain 2007). Ultimately, the NIWC's sustained, twofold efforts, both at the negotiation table and in the community, provided the essential foundation of inclusion and comprehensiveness that continues to underwrite the durability of peace in Northern Ireland.

Cultural Support as a Prerequisite for Women's Inclusion in Peace Processes

Despite the strength of these case studies, critics argue that women's inclusion is not feasible in societies lacking cultural support for gender equality. They contend that attempting to "engineer" inclusion risks delegitimizing peace processes or creating backlash that could destabilize fragile political orders (Demirci et al. 2023). Critics point to a 2016 Colombian case study when a slim majority rejected the initial peace agreement draft in a referendum. A key factor in the opposition was the no-campaign,

which fueled concerns that the agreement's focus on "gender ideology" would undermine traditional family structures (Bramsen 2022). While Congress ultimately ratified a revised treaty, the popular rejection of the first version clearly demonstrates the difficult trade-off that exists between pursuing peace and promoting inclusion. According to this argument, gender inclusion must follow cultural change rather than precede it; in deeply patriarchal societies, inserting women into negotiations is seen as too radical a departure from established social norms.

The critique that cultural support must be a prerequisite for female inclusion in peacebuilding, however, fundamentally misinterprets both the empirical evidence and the causal mechanisms of peacebuilding. First, the claim that gender inclusion destabilizes negotiations lacks support. A landmark study by Krause and Bränfors found that, when women's groups were included in peace processes, agreements were 64 percent less likely to fail and 35 percent more likely to last at least fifteen years (Krause, Krause, and Bränfors 2018). This relationship holds even in patriarchal or religiously conservative contexts. Second, many scholars believe the examples critics invoke are often misinterpreted (Bramsen 2022). Colombia's failed 2016 referendum was not caused by women's participation, but rather by public opposition to the lenient terms offered to Revolutionary Armed Forces of Colombia—People's Army (FARC) rebels, a strong "No" campaign led by former President Álvaro Uribe, low voter turnout, and hurricane Matthew, which kept the northern coastal region who was largely in support of the "yes" campaign from mobilizing (Klobucista 2017). The argument also overlooks the evidence that women are influential peacebuilders in culturally conservative societies, including Liberia, Rwanda, and the Philippines (Chang et al. 2015). Cultural resistance did not prevent women from shaping durable peace in these cases; rather, peace processes served as catalysts for broader social transformation.

By excluding women, countries reinforce patriarchal norms and miss an opportunity to broaden the base of peace support. Women's organizations often represent constituencies

otherwise absent from negotiation rooms, providing information about community needs, early warning indicators, and the root causes of conflict, information peacebuilders need to produce durable agreements. Fundamentally, the argument for keeping women on the sidelines mistakenly assumes that society must change its culture before women can be included. Peace processes are moments of political rupture in which institutional, legal, and social norms are renegotiated. As Ní Aoláin argues, post-conflict transitions are precisely the moment when gender norms *can* change because the political order is reconstructed (Ní Aoláin 2011).¹ Waiting for cultural change *before* including women in the peace process reverses this logic: inclusion is not the conclusion of transformation...it is one mechanism that drives it.

Peacebuilders who believe gender inclusion is 'too disruptive' misunderstand that *peace itself is disruptive*. Peace restructures power. Peace reallocates resources. Peace transforms governance. The inclusion of women does not destabilize this peace. On the contrary, women strengthen peace by enhancing its legitimacy and ensuring its comprehensiveness.

Conclusions and Implications

While the importance of gender inclusion is now well established in peacebuilding scholarship and policy frameworks, a persistent implementation gap remains between rhetorical commitment and actual practice. The central challenge is no longer demonstrating that women's participation contributes to more durable peace, but ensuring that this insight is systematically translated into the design and execution of peace processes. At the same time, there are growing indications that the international community may be moving away from the gender mainstreaming necessary for sustainable peace. Shifting geopolitical priorities and declining institutional emphasis on inclusion risk reinforcing a model of peacebuilding that privileges short-term elite bargains over long-term societal transformation. This makes the task of institutionalizing gender inclusion more urgent, not less.

Theories of peace durability highlight legitimacy, monitoring, comprehensiveness,

1 Ní Aoláin and D Fionnuala, "Advancing Women's Rights in Conflict and Post-Conflict Situations," *SSRN Electronic Journal*, February 23, 2011, <https://papers.ssrn.com/sol3/Delivery.cfm?abstractid=1769013>.

and credible commitments as key mechanisms underpinning sustainable agreements. Existing research and case evidence suggest that women's participation strengthens each of these dimensions. However, this relationship remains inconsistently reflected in practice. Moving beyond symbolic inclusion requires formal mechanisms such as gender quotas in negotiation delegations, dedicated Gender Sub-Commissions, and sustained support for women's civil society networks. These measures embed participation within institutional structures rather than leaving it contingent on political will.

Moreover, broadening the scope of peacebuilding beyond military and territorial concerns remains essential. Women's contributions have consistently expanded peace agendas to include social and economic dimensions such as education, healthcare, trauma recovery, and livelihoods—issues closely tied to positive peace and long-term stability. Yet, as illustrated in cases like Colombia, the inclusion of such provisions is only meaningful when matched by adequate resources, monitoring, and implementation capacity.

If gender inclusion is a core predictor of durable peace, then peacebuilders who treat it as peripheral undermine their own objectives. Taken together, the evidence reinforces a clear but insufficiently realized insight: durable peace depends not only on negotiated settlements among elites, but on the extent to which those settlements reflect and engage the broader society. Women's participation plays a critical role in bridging this gap by enhancing legitimacy, widening the substantive scope of agreements, and strengthening their social foundations. The implications are straightforward. Treating gender inclusion as peripheral is not simply a normative shortcoming—it undermines the effectiveness of peacebuilding itself. If current trends continue to deprioritize gender mainstreaming, the gap between what is known to support durable peace and what is practiced may widen further. For societies recovering from conflict, the question is not whether they can “afford” to include women in peacebuilding; it is whether they can afford not to.

References

- Bouvier, Virginia M., and UN Women. 2016. "Gender and the Role of Women in Colombia's Peace Process." Background paper. March 4, 2016. <https://www.usip.org/sites/default/files/Gender-and-the-Role-of-Women-in-Colombia-s-Peace-Process-English.pdf>.
- Bramsen, Isabel. 2022. "Controversies of Inclusion in the Colombian Peace Process: The Balancing Act of Introducing New Norms and Gaining Popular Support." Policy Brief. Oslo: Peace Research Institute Oslo (PRIO). <https://reliefweb.int/report/colombia/controversies-inclusion-colombian-peace-process-balancing-act-introducing-new-norms>.
- Caprioli, Mary. 2000. "Gendered Conflict." *Journal of Peace Research* 37, no. 1 (January): 51–68. <http://www.jstor.org/stable/425725>.
- Center for Security Studies and Development (CESSD). 2024. "Understanding the Barriers to Women's Participation in Liberia's Reconciliation and Peace Building Processes." Report. Genius Solutions Group. <https://globaldemocracycoalition.org/wp-content/uploads/2024/10/FFINaLLSURVEYREPORT.pdf>.
- Chang, Patty, Mayesha Alam, Roslyn Warren, Rachel Vogelstein, and Natalie Walker. 2015. *Women Leading Peace: A Close Examination of Women's Political Participation in Peace Processes in Northern Ireland, Guatemala, Kenya, and the Philippines*. Washington, DC: Georgetown Institute for Women, Peace and Security. <https://giwps.georgetown.edu/wp-content/uploads/2017/10/Women-Leading-Peace.pdf>.
- Cockburn, Cynthia. 2007. *From Where We Stand: War, Women's Activism and Feminist Analysis*. London: Zed Books. <https://doi.org/10.5040/9781350220287>.
- Demirci, Irem, Simone Wisotzki, and Clara-Auguste Sophie Mueller. 2023. "Backlash Against and Resistance to Feminist Peacebuilding." PRIF Reports no. 6. Frankfurt: Leibniz Institute for the Social Sciences (GESIS). <https://doi.org/10.48809/prifrep2306>.
- Doyle, Michael W., and Nicholas Sambanis. 2011. *Making War and Building Peace*. Princeton: Princeton University Press eBooks. <https://doi.org/10.2307/j.ctt7rtn4>.
- Fortna, Virginia Page. 2004. "Peace Mechanisms: What Works and What Doesn't?" In *Peace Time: Cease-Fire Agreements and the Durability of Peace*, 173–210. Princeton: Princeton University Press. <http://www.jstor.org/stable/j.ctv39x5v5>.
- Galtung, Johan. 1969. "Violence, Peace, and Peace Research." *Journal of Peace Research* 6, no. 3: 167–91. <https://doi.org/10.1177/002234336900600301>.
- Hain, Peter. 2007. "Peacemaking in Northern Ireland: A Model for Conflict Resolution?" Speech, United Kingdom of Great Britain and Northern Ireland, June 12. <https://cain.ulster.ac.uk/issues/politics/docs/nio/ph120607.pdf>.
- Joshi, Madhav, and Jason Michael Quinn. 2017. "Implementing the Peace: The Aggregate Implementation of Comprehensive Peace Agreements and Peace Duration After Intrastate Armed Conflict." *British Journal of Political Science* 47, no. 4 (October): 869–92. <https://doi.org/10.1017/s0007123415000381>.
- Kieh, George Klay. 2009. "The Roots of the Second Liberian Civil War." *International Journal on World Peace* 26, no. 1 (March): 7–31. [suspicious link removed].
- Kieh, George Klay. 2015. "Ending Liberia's Second Civil War: Religious Women as Peacemakers." Berkley Center for Religion, Peace, and World Affairs. September 25, 2015. <https://repository.digital.georgetown.edu/handle/10822/1051849>.
- Kilmurray, Avila, and Monica McWilliams. 2012. "Struggling for Peace: How Women in Northern Ireland Challenged the Status Quo." *SSRN Electronic Journal* (February 1). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2489950.

- Kirby, Dianne. 2018. "Women of Faith and the Northern Ireland Peace Process: Breaking the Silence." *openDemocracy*. April 10, 2018. <https://www.opendemocracy.net/en/transformation/women-of-faith-and-northern-ireland-peace-process-breaking-silence/>.
- Klobucista, Claire. 2017. "Colombia's Civil Conflict." Council on Foreign Relations. January 11, 2017. <https://www.cfr.org/backgrounder/colombias-civil-conflict>.
- Krause, Jana, Werner Krause, and Piia Bränfors. 2018. "Women's Participation in Peace Negotiations and the Durability of Peace." *International Interactions* 44, no. 6: 985–1016. <https://doi.org/10.1080/03050629.2018.1492386>.
- Maran, Emily. 2025. "Beyond the Glass Ceiling: Ellen Johnson Sirleaf's Presidency and the Fight for Women's Political Inclusion in Liberia." *New Century Scholar* (May 31). <https://journals.tulane.edu/ncs/article/view/4043>.
- Matanock, Aila M., and Natalia Garbiras-Díaz. 2018. "Considering Concessions: A Survey Experiment on the Colombian Peace Process." *Conflict Management and Peace Science* 35, no. 6 (August 6): 637–55. <https://doi.org/10.1177/0738894218787784>.
- Ní Aoláin, Fionnuala. 2011. "Advancing Women's Rights in Conflict and Post-Conflict Situations." *SSRN Electronic Journal* (February 23). <https://papers.ssrn.com/sol3/papers.cfm?abstractid=1769013>.
- Puechguirbal, Nadine. 2012. "The Cost of Ignoring Gender in Conflict and Post-Conflict Situations: A Feminist Perspective." *Amsterdam Law Forum* 4, no. 1 (Winter): 4–19. <https://doi.org/10.37974/alf.207>.
- Quintana, Ana Rosa. 2016. "The Colombian–FARC Peace Deal: Why It Failed, and How the U.S. Can Support a Responsible Renegotiation." Report. Washington, DC: The Heritage Foundation. October 24, 2016. <https://www.heritage.org/global-politics/report/the-colombian-farc-peace-deal-why-it-failed-and-how-the-us-can-support>.
- Tripp, Aili Mari. 2015. *Women and Power in Postconflict Africa*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/cbo9781316336014>.
- UN Office of the Special Adviser on Gender Issues and Advancement of Women (OSAGI). 2000. "Landmark Resolution on Women, Peace and Security (Security Council Resolution 1325)." October 31, 2000. <https://www.un.org/womenwatch/osagi/wps/>.
- UN Women. 2015. *A Global Study on the Implementation of United Nations Security Council Resolution 1325*. New York: UN Women. <https://unwomen.de/wp-content/uploads/2022/03/UNW-GLOBAL-STUDY-1325-2015.pdf>.
- UN Women. 2025. "The Peace Deal That Put Women First: What Colombia Taught the World." July 10, 2025. <https://www.unwomen.org/en/news-stories/feature-story/2025/07/the-peace-deal-that-put-women-first-what-colombia-taught-the-world>.