Chapman Law Review

CHAPMAN UNIVERSITY ADMINISTRATION

DANIELE C. STRUPPA, PH.D.
President

GLENN M. PFEIFFER, PH.D.
Executive Vice President, Provost and Chief Academic Officer

BRIAN THOMASON
Executive Vice President and Chief Operating Officer

MATTHEW PARLOW
Executive Vice President and Chief Advancement Officer

BRENDA RUSHFORTH
Vice President and Chief Human Resources Officer

COLLETTE CREPPELL
Vice President of Campus Planning and Design

HELEN NORRIS
Vice President and Chief Information Officer

MARIE WILLIAMS
Vice President of Marketing and Communications

MARTINA NIESWANDT, PH.D
Vice President of Research

JANNA BERSI, ED.D.
Senior Vice President, Investments and Administration

JERRY PRICE, PH.D.
Senior Vice President for Student Affairs and Dean of Students

MIKE PELLY
Vice President and Dean of Enrollment

HIMA VATTI
Senior Vice President of Legal Affairs and General Counsel

GABRIELA CASTAÑEDA, ED.D.
Vice President of First-Generation and Promising Futures Program

AMY ROGAN-MEHTA, ESQ.
Senior Vice President of University Advancement

BOARD OF TRUSTEES

PARKER S. KENNEDY, *Chair*
JAMES V. MAZZO, *Vice Chair*
SCOTT CHAPMAN, *Secretary*
MARILYN ALEXANDER
LISA ARGYROS '07
RAJ S. BHATHAL
GAYE BIRTCHER
KEN BUNT '93
ROBERT CASE
AKIN CEYLAN '90
IRVING M. CHASE
JEROME W. CWIERTNIA
ZEINAB DABBAH (JD '12)
DALE E. FOWLER '58
GAVIN S. HERBERT, JR.
MARK HILBERT
ANDY HOROWITZ
ANDRÉS IRLANDO
MARK CHAPIN JOHNSON '05
JASON KROTTS '00
DANIELLE LIMP
RICK LOWE
MELINDA MASSON
MARONYA MOULTRIE
MARYBELLE MUSCO
RICHARD MUTH (MBA '81)
JAMES B. ROSZAK
THE HONORABLE LORETTA SANCHEZ '82
MOHINDAR S. SANDHU

KEVIN SCHEETZ
RONALD M. SIMON
MAX VALDES
TIM VANDERHOOK
THE HONORABLE GADDI H. VASQUEZ '09
ANNETTE WALKER
GEORGE WALL
NELLA WEBSTER O'GRADY '71
KAREN R. WILKINSON '69
JANE FUJISHIGE YADA
CHARLIE ZHANG

EMERITUS CHAIRS

WYLIE A. AITKEN
THE HONORABLE GEORGE L. ARGYROS '59
DOY B. HENLEY
PARKER S. KENNEDY
DONALD E. SODARO

EMERITUS TRUSTEES

DONNA FORD ATTALLAH '61
ARLENE R. CRAIG
DAVID C. HENLEY
ROGER C. HOBBS
WILLIAM K. HOOD
JOANN LEATHERBY
TOM MALLOY
RICHARD R. SCHMID
R. DAVID THRESHIE

EX-OFFICIO TRUSTEES

REVEREND LATAUNYA BYNUM '76
KATHY DUNLAP
LISA HALLAIAN
REVEREND JAY HARTLEY
REVEREND DAYNA KINKADE
MICHAEL PENN (JD '04)
MICHELE PHILO '03
REVEREND RACHAEL PRYOR
REVEREND RICHIE SANCHEZ
DANIELE C. STRUPPA

BOARD OF ADVISORS

THOMAS D. PHELPS, *Chair*
WYLIE A. AITKEN
ROBERT ALVARADO
PHILLIP H. CASE
ROBERT E. CURRIE
JOHN R. EVANS
WOLFGANG FRISCH '97 (JD '00)
THE HONORABLE RICHARD D. FYBEL
THE HONORABLE ANDREW J. GUILFORD
DOY B. HENLEY
JANET E. HUMPHREY
PARKER S. KENNEDY
THE HONORABLE LAYNE H. MELZER
DAVID MURPHY
PAUL D. PATON

SCHOOL OF LAW ADMINISTRATION

PAUL D. PATON, JSD
Dean and Donald P. Kennedy Chair in Law

RICHARD E. REDDING
Associate Dean for Academic Affairs and Professor of Legal Research and Writing

DEEPA BADRINARAYANA
Associate Dean for Research and Faculty Development

MARIO MAINERO
Associate Dean for Governance and Strategic Initiatives & Bar Preparation and Academic Standards

JUSTIN CRUZ
Associate Dean of Admission and Diversity Initiatives

SEAN BIGLEY
Senior Assistant Dean of Administration

CAMILLE HEENAN
Assistant Dean for Career Services

SARIRA A. SADEGHI
Sam & Ash Assistant Dean for Academic Achievement

MARYAM ISLES
Assistant Dean for Academic Services and Registration

JESSICA JOHN
Assistant Dean for Student Affairs

LAW SCHOOL FACULTY

DEEPA BADRINARAYANA
Professor of Law

MICHAEL BAZYLER
Professor of Law
1939 Society Law Scholar in Holocaust and Human Rights Studies

TOM W. BELL
Professor of Law

DENIS BINDER
Professor of Law

JOHN BISHOP
Professor of Law
Visiting Faculty

DANIEL BOGART
Professor of Law
Bolinger Chair in Real Estate, Land Use and Environmental Law

DR. THOMAS CAMPBELL
Professor of Law
Doy and Dee Henley Distinguished Professor of Jurisprudence
Former Dean (2011–2016)

LAN CAO
Professor of Law
Betty Hutton Williams Professor of International Economic Law
Director of International Law Emphasis Program

JENNY CAREY
Professor of Law
Professor of Legal Research and Writing

BOBBY DEXTER
Professor of Law

KURT EGGERT
Professor of Law
Director of the Alona Cortese Elder Law Center

JUDD FUNK
Professor of the Practice of Entertainment Law

DR. JOHN A. HALL
Professor of Law

ERNESTO HERNANDEZ
Professor of Law

HUGH HEWITT
Professor of Law

SCOTT W. HOWE
Frank L. Williams Professor of Criminal Law
Former Interim Dean (2010–2011, 2016)

NAHAL KAZEMI
Assistant Professor

JANINE KIM
Wylie A. Aitken Professor of Law, Race, and Social Justice

CAROLYN LARMORE
Professor of the Practice of Law
Director of the Externship Program

STEPHANIE LASCELLES
Associate Professor of Legal Research and Writing

MARIO MAINERO
Professor of Academic Achievement and Bar Services
The Gray Family Professor of Law
Associate Dean for Bar Preparation and Academic Achievement

CELESTINE RICHARDS MCCONVILLE
Professor of Law
Henry Salvatori Professor of Law and Community Service

HENRY NOYES
Professor of Law

MATTHEW J. PARLOW
Executive Vice President/CAO
Parker S. Kennedy Professor of Law

PAUL D. PATON, JSD
Dean
Donald P. Kennedy Chair in Law & Ethics

ABIGAIL PATTHOFF
Professor of Legal Research and Writing

DR. RICHARD REDDING
Professor of Law
Ronald D. Rotunda Distinguished Professor of Jurisprudence
Associate Dean for Academic Affairs

SUSANNA K. RIPKEN
Professor of Law
William P. Foley II Chair in Corporate Law and Taxation

LAWRENCE ROSENTHAL
Professor of Law

MARY LEE RYAN
Professor of the Practice of Entertainment Law
Director of Entertainment Law Program

DR. VERNON L. SMITH
George L. Argyros Endowed Chair in Finance and Economics
Professor of Economics and Law
Smith Institute for Political Economy and Philosophy

KENNETH STAHL
Professor of Law
Director of the Environmental Land Use and Real Estate Law Program

DR. WILLIAM STALLWORTH
Professor Emeritus of Law

KIM TYLER
Professor of Law
Visiting Faculty

MATTHEW TYMANN
Assistant Professor of Law

DR. PARHAM H. WILLIAMS, JR.
Professor Emeritus
Former Dean (1997–2007)

GEORGE WILLIS
Professor of Law
Clinical Faculty
Director of the Tax Law Clinic
Director of the Tax Law Program

DR. BART J. WILSON
Professor of Law
Donald P. Kennedy Chair in Economics and Law
Director of the Smith Institute for Political Economy and Philosophy

Chapman Law Review

Volume 28 Spring 2025 Number 2

ARTICLES

NOTE

Editor's Note

As Chapman University Dale E. Fowler School of Law celebrates its thirtieth anniversary, *Chapman Law Review* commemorates this milestone by expanding our publication efforts for Volume 28. This year, we are proud to present three issues—two general law review issues and one Symposium Issue—each centered around a shared editorial mission: to feature legal professionals from all walks of life, explore a multiplicity of legal perspectives, and broaden the reach of legal scholarship beyond the confines of the legal profession. The articles published in this issue reflect the richness of that goal and represent the thoughtful, rigorous, and deeply relevant work of scholars, practitioners, and students alike.

This Volume also reflects our effort to highlight the voices and values of our student body. Throughout the year, we invited our members to assist in identifying articles that resonate with our community and reflect the issues that matter most to our generation of legal professionals. I would especially like to thank our Executive Article Submissions Editors, Tyler Cruz and Allison Lloyd, for their continued dedication to this mission. Their insight and persistence helped ensure that the scholarship we feature speaks not only to the legal field, but to the lived realities and aspirations of our peers.

Professor Stephen Hendricks opens the issue with an article that provides a nuanced critique of the Federal Trade Commission's 2024 nationwide ban on non-compete agreements. While recognizing the rule's purpose to protect worker mobility, Hendricks argues that a categorical prohibition oversimplifies the balance between labor freedom and employer interests. Drawing on Germany's *Karenzentschädigung* model, he proposes a federal compensatory framework in which non-competes are enforceable only if employers provide post-employment pay. His article outlines a constitutionally sound middle path that promotes innovation while ensuring economic fairness.

In his article, Arvid Kerschnitzki re-examines the First Amendment in light of the challenges posed by the digital age. While courts have long protected speech in the name of individual liberty and open discourse, Kerschnitzki argues that this framework has failed to account for the ways social media distorts public dialogue—spreading misinformation, reinforcing echo chambers, and undermining democratic deliberation. Drawing on the theory of militant democracy, the article contends that the First Amendment must be recalibrated to not

only preserve liberty but also to protect the democratic process itself by promoting speech that is inclusive, rational, and conducive to an informed citizenry.

Professor Paul Rogerson's article follows and explores the cyclical nature of U.S. patent law, where periods of strong patent protections are followed by eras of weakening rights. While past research has qualitatively linked this pattern to waves of technological change, Rogerson introduces a new, quantitative approach to test this theory over two centuries. Using historical productivity data and patent records, he has built a model that shows courts consistently tighten or loosen patent standards in response to technological surges—typically with a decade-long lag. His findings confirm that the legal system reacts predictably to innovation, echoing patterns from the 1800s to today.

In the final article, Evan Yahng addresses a growing legal challenge in the age of data breaches and evolving technology: what qualifies as "unreasonable delay" under state data breach notification laws. While every U.S. state and territory requires companies to notify individuals of compromised data, courts have long avoided defining delay due to early dismissals in litigation. However, in 2024, that all changed. Yahng reviews recent decisions where courts relied solely on the number of days that had elapsed to find delays unreasonable. He argues that this approach misinterprets the purpose of notification statutes, and it risks both over-penalizing companies and harming consumers. Instead, he proposes a standard that focuses on the quality of the company's post-breach investigation.

The issue ends with a student note written by Ms. Aubrey Butler, a J.D. Candidate currently in her third year of study at Chapman University Dale E. Fowler School of Law, Class of 2025. During her second year, Aubrey served as a Staff Editor for *Chapman Law Review*. She has held the critical position of Managing Editor during her third year and, in that capacity, she has been instrumental in the production and publication of this Volume. Aubrey's note examines the growing use of artificial intelligence and machine-generated evidence in the courtroom, with a focus on software tools like TrueAllele and COMPAS. Highlighting the dangers of opaque algorithms and proprietary "black box" systems, she proposes the creation of a new federal agency with a dedicated department to oversee and advise courts on AI-driven evidence, ensuring scientific integrity and due process protections across the legal system. In recognition of the strength and relevance of her work, the *Chapman Law Review*

Faculty Committee nominated Aubrey's note for the prestigious Scribes Award.

Chapman Law Review is deeply grateful for the unwavering support of the faculty, administration, and library staff who made the publication of this issue possible. We extend our sincerest thanks to our extraordinary faculty advisor, Professor Celestine McConville, whose wisdom, encouragement, and steadfast guidance have shaped every stage of this Volume. Her mentorship has been a constant source of strength and clarity throughout the year. We are also profoundly grateful to Dean Paul D. Paton for his leadership and enthusiasm, and to our dedicated Faculty Advisory Committee—Professors Kurt Eggert, Janine Kim, Carolyn Larmore, Lawrence Rosenthal, and Matthew Tymann—for their thoughtful insight and continued investment in our success. Finally, we owe special thanks to the brilliant Research Librarians of the Hugh & Hazel Darling Library, Heather Agnew and Phillip Der Mugrdechian, whose tireless support and deep expertise have been indispensable to our editorial process. This Volume would not exist without each of you, and we are truly honored to have had your partnership.

I would like to especially acknowledge the incredible efforts of our Executive Managing Editor, Anna Ross, and our Executive Production Editor, Sara Moradi. Their tireless dedication, adaptability, and hard work were fundamental to the success of this Volume. Last but certainly not least, I would like to thank the staff of the 2024–2025 *Chapman Law Review*. Your remarkable, committed, and indefatigable work was paramount to the publication of this Volume. I am truly honored and privileged to have been part of this journal and for the opportunity to serve and lead *Chapman Law Review* over this past year.

Serving as Editor-in-Chief has been the most fulfilling experience of my law school career. It has challenged me as a leader, strengthened me as a writer and editor, and reminded me daily of the value of collaboration, mentorship, and service. I am endlessly grateful for the friendships formed, the lessons learned, and the trust placed in me to guide this team through such a meaningful year. Thank you for allowing me the opportunity to help shape this Volume—it has been the greatest privilege.

"*Fight for the things that you care about, but do it in a way that will lead others to join you.*"

— Justice Ruth Bader Ginsburg

As we reflect on the legacy represented in these pages, Justice Ginsburg's words are a powerful reminder of what it means to be part of this profession—not only to care deeply about justice, but to advocate in ways that bring others along with us. That spirit of courage, collaboration, and conviction defines this Volume, and it is the legacy we hope to carry forward into the future.

Taline Nicole Ratanjee
Editor-in-Chief

Breaking the Bind: Rethinking Non-Compete Agreements in a Federal Framework

Stephen M. Hendricks

Breaking the Bind: Rethinking Non-Compete Agreements in a Federal Framework

Stephen M. Hendricks[*]

In April 2024, the Federal Trade Commission (FTC) adopted a sweeping rule that bans most non-compete agreements nationwide. Though well-intentioned, the FTC's categorical prohibition neglects the complex interplay between worker mobility, employer investment, and the protection of legitimate business interests. This Article contends that the United States must move beyond the current false binary of complete prohibition or unregulated enforcement. Drawing on comparative legal analysis, it proposes a federal framework modeled on Germany's Karenzentschädigung system, under which employers must provide post-employment compensation—typically fifty percent of prior salary—to enforce non-competes.

After tracing the evolution of non-compete doctrine in American law, this Article critiques the FTC's rule on constitutional, statutory, and federalism grounds. It then offers a detailed roadmap for legislative reform, demonstrating how a structured compensatory approach can promote innovation, protect trade secrets, and ensure economic security for workers. The proposal includes statutory minimums on compensation and duration, administrative enforcement through the Department of Labor, and judicial review grounded in proportionality and business necessity. Empirical evidence from Germany supports the model's efficacy in reducing litigation, enhancing labor mobility, and preserving competitive markets.

Ultimately, this Article advances a middle path—"earned enforceability"—that reconciles economic dynamism with legal stability. By transforming non-compete agreements from instruments of suppression into deliberate, reciprocal contracts, a federal compensatory regime can replace the existing regulatory patchwork with a durable, constitutionally sound solution for the knowledge-based economy.

* Stephen Hendricks is an Assistant Professor at Western State College of Law, where he teaches Contracts and Business Associations. He also directs the Western State College of Law Entrepreneurship Clinic. A special thank you to Associate Dean for Research & Faculty Development Stacey Sobel for her valuable guidance and support. Thank you to Dean Marisa Cianciarulo and Associate Dean for Academic Affairs Paul Arshagouni for their valuable guidance and support. Thank you to the many professors at Western State College of Law for their support and valuable comments on earlier drafts. Many thanks to my outstanding research assistant Amanat Kular.

I. INTRODUCTION

On May 7, 2024, the Federal Trade Commission (FTC) took unprecedented action by issuing a final rule that effectively bans most non-compete agreements across the United States.[1] This sweeping rule, which applies to nearly all employers and workers, represents the most significant federal intervention into employment contract law in recent history.[2] While the FTC's rule aims to promote worker mobility and market competition, it fails to recognize the legitimate business interests that non-compete agreements have historically protected.[3] This binary approach—complete prohibition rather than measured regulation—threatens to disrupt established business practices and potentially harms the very innovation economy it seeks to protect.

The rule has been met with several legal challenges, with opponents arguing that the FTC exceeded its statutory authority and that the rule represents an unconstitutional exercise of federal power.[4] In *Ryan, LLC v. FTC*, one such challenge resulted in a nationwide injunction.[5] The FTC initiated an appeal on October 18, 2024,[6] but following a change in presidential administration and leadership at the agency,[7] it requested a 120-day pause in the appellate proceedings.[8] The FTC is expected

---

1 FTC Non-Compete Clause Rule, 16 C.F.R. §§ 910.1–.6 (2024); *see also Open Commission Meeting — April 23, 2024*, FED. TRADE COMM'N (Apr. 23, 2024, 2:00 PM), https://www.ftc.gov/news-events/events/2024/04/open-commission-meeting-april-23-2024 [https://perma.cc/Y75W-UBBP].

2 *See* Tom Shumate, *Game Over or Game On? The Future of Noncompetes*, 60 TENN. BAR J. 18, 22 (2024).

3 *See* Michael K. Molzberger, *The Proposed Nationwide Ban on Non-Competition Agreements by the Federal Trade Commission*, 43 FRANCHISE L.J. 1, 2 (2024).

4 *See* Bruce Allain, *FTC's Non-Compete Ban Hit with Multiple Legal Challenges*, THE SOURCE ON HEALTHCARE PRICE & COMPETITION (May 15, 2024), https://sourceonhealthcare.org/ftcs-non-compete-ban-hit-with-multiple-legal-challenges/ [https://perma.cc/HW55-Y58A].

5 Ryan, LLC v. FTC, 746 F. Supp. 3d 369, 387, 389–90 (N.D. Tex. 2024).

6 *See* Notice of Appeal, Ryan, LLC v. FTC, 746 F. Supp. 3d 369 (N.D. Tex. 2024) (No. 24-10951).

7 *See* Jesse M. Coleman & Eron Reid, *FTC Requests Stay of Appeals to Challenges to FTC Non-Compete Rule Citing New Administration*, SEYFARTH (Mar. 11, 2025), https://www.tradesecretslaw.com/2025/03/articles/ftcs-crackdown-on-non-competes/ftc-requests-stay-of-appeals-to-challenges-to-ftc-non-compete-rule-citing-new-administration/ [https://perma.cc/QUZ2-N8Q7] ("Before ascending to lead the agency, then-Commissioner [Andrew] Ferguson opposed the Rule, arguing that it lacked the authority for broad rulemaking to ban non-compete agreements while also offering pro-business justifications against the ban.").

8 Cody D. Woods, *FTC Forms Labor Task Force; Noncompete Appeals Stayed*, THE NAT'L L. REV. (May 1, 2025), https://natlawreview.com/article/ftc-forms-labor-task-force-noncompete-appeals-stayed [https://perma.cc/EY7S-L7CS].

to inform the court by mid-July 2025 whether it intends to continue pursuing the appeal.[9] Even if the FTC's rule is abandoned, the legal uncertainty surrounding non-compete agreements will persist, causing courts to grapple with their scope and enforceability, necessitating a solution.

Non-compete agreements have long served as crucial tools for businesses to protect their intellectual property, customer relationships, and investments in employee training.[10] When properly structured, these agreements can foster innovation by giving employers the confidence to share trade secrets and invest in employee development.[11] However, their misuse through overly broad restrictions or application to low-wage workers has rightfully drawn criticism and scrutiny.[12] The difficulty, therefore, lies not in wholesale prohibition, but in crafting a balanced framework that protects both legitimate business interests and worker rights.

This Article proposes a solution: rather than maintaining the current state-by-state patchwork or pushing forward with the FTC's complete ban, Congress should enact federal legislation modeled on Germany's sophisticated approach to non-compete agreements. The German system, known as the *Karenzentschädigung* model, requires employers to provide substantial compensation, typically fifty percent of an employee's total remuneration, during the non-compete period.[13] This approach has successfully balanced employer and employee interests for decades while maintaining a dynamic and competitive economy.[14]

The United States can learn from this model while adapting it to American legal and business contexts. A federal framework

---

9 *See id.*

10 Norman D. Bishara, *Covenants Not to Compete in a Knowledge Economy: Balancing Innovation from Employee Mobility Against Legal Protection for Human Capital Investment*, 27 BERKELEY J. EMP. & LAB. L. 287, 294–96 (2006).

11 Chris Jackson & Jason Wiens, *A Fair Fight: Entrepreneurship and Competition Policy*, EWING MARION KAUFFMAN FOUND. (July 8, 2016), https://www.kauffman.org/resources/entrepreneurship-policy-digest/a-fair-fight-entrepreneurship-and-competition-policy/ [https://perma.cc/QC7U-BA7F].

12 *See* Matt Marx & Lee Fleming, *Non-Compete Agreements: Barriers to Entry . . . and Exit?*,12 INNOVATION POL'Y & ECON. 39, 39, 45–61 (2012).

13 *See generally* WOLFGANG HROMADKA & FRANK MASCHMANN, ARBEITSRECHT BAND 1 § 10, at 389–556 (6th ed. 2015) (discussing the German statutory requirements for non-compete agreements).

14 *See* Hagen Köckeritz & Guido Zeppenfeld, *Germany: Restrictive Covenants*, MAYER BROWN (July 25, 2024), https://www.mayerbrown.com/en/insights/publications/2024/07/restrictive-covenants-germany [https://perma.cc/UW7D-W5WC].

based on the German system would bring clarity and consistency to non-compete enforcement while ensuring fair treatment of workers. Such legislation would need to address several key challenges: (1) establishing appropriate compensation levels, (2) defining reasonable duration limits, (3) creating effective enforcement mechanisms, and (4) navigating complex constitutional issues regarding federal authority in traditional state law domains.

This Article proceeds in several parts. Part II examines the historical background of non-compete agreements in American law and analyzes the current state-by-state approach, culminating in a discussion of the FTC's new rule. Part III addresses the constitutional implications of federal regulation in this area, particularly focusing on Commerce Clause authority and potential challenges to both the FTC rule and proposed legislation. Part IV provides a comprehensive analysis of the German model, examining its key features and implementation success. Part V takes an economic and empirical look at non-competes, while Part VI presents a detailed proposal for a federal framework, adapting the German approach to the American context while addressing unique U.S. legal and business considerations. Part VII contemplates the implementation and policy considerations surrounding such a framework. Part VIII anticipates and addresses potential criticisms of the proposed framework, and Part IX concludes.

The time has come for the United States to move beyond the binary choice between unrestricted non-compete agreements and their outright prohibition. By examining successful international models and crafting thoughtful federal legislation, the United States can create a framework that protects innovation, promotes fair competition, and respects worker rights. This Article provides a roadmap for achieving that balance.

II. HISTORICAL BACKGROUND AND CURRENT STATE OF NON-COMPETE LAW

A. Evolution of Non-Compete Agreements in American Law

The history of non-compete agreements in American jurisprudence traces back to the English common law, where courts initially viewed such restrictions with significant skepticism. The seminal 1711 case of *Mitchel v. Reynolds* established the foundation for modern non-compete doctrine by introducing a "rule of reason" approach that balanced an

employer's protectable interests against public policy concerns about restraints on trade.[15] This framework crossed the Atlantic and took root in American courts, though its application evolved significantly as the American economy transformed from agrarian to industrial to information-based.[16]

Early American courts generally disfavored non-compete agreements, viewing them primarily through the lens of economic liberty and free market principles. However, as businesses began investing more heavily in trade secrets, customer relationships, and employee training during the Industrial Revolution, courts gradually recognized legitimate business interests that could justify reasonable restrictions on post-employment competition.[17] This shift reflected the growing complexity of employer-employee relationships and the increasing importance of intellectual property protection in the American economy.

The twentieth century saw further refinement of non-compete doctrine as courts developed more sophisticated tests for evaluating these agreements. The modern approach typically examines three key elements: (1) the existence of legitimate business interests worthy of protection, (2) the reasonableness of the restrictions in terms of geographic scope and duration, and (3) the public interest in maintaining competition and employee mobility.[18] This framework, while widely adopted, has been applied with varying degrees of stringency across jurisdictions, leading to the current patchwork of state approaches.

B. Current State Approaches

The regulation of non-compete agreements has traditionally fallen within state jurisdiction, resulting in significant variation across the United States. This diversity of approaches reflects different policy priorities and economic philosophies among states, creating a complex landscape for both employers and employees to navigate.[19]

---

15 Mitchel v. Reynolds (1711) 24 Eng. Rep. 347, 348–49 (Gr. Brit.).

16 *See* Catherine L. Fisk, *Working Knowledge: Trade Secrets, Restrictive Covenants in Employment, and the Rise of Corporate Intellectual Property, 1800-1920*, 52 HASTINGS L.J. 441, 442, 534–35 (2001).

17 1 PETER S. MENELL ET AL., INTELLECTUAL PROPERTY IN THE NEW TECHNOLOGICAL AGE: 2020, at 115–17, 120–21 (2020).

18 *See* RESTATEMENT (SECOND) OF CONTS. § 188 (AM. L. INST. 1981).

19 *See* Norman D. Bishara, *Fifty Ways to Leave Your Employer: Relative Enforcement of Covenants Not to Compete, Trends, and Implications for Employee Mobility Policy*, 13 U. PA. J. BUS. L. 751, 753, 756 (2011).

1. The California Model: Complete Prohibition

California stands as the prime example of a jurisdiction that has taken an absolutist position against non-compete agreements. Since 1872, section 16600 of the California Business and Professions Code has provided that "every contract by which anyone is restrained from engaging in a lawful profession, trade, or business of any kind is to that extent void."[20] California courts have interpreted this statute broadly, consistently striking down non-compete agreements except in very limited circumstances involving the sale of businesses or dissolution of partnerships.[21]

Proponents of the California approach argue that its prohibition on non-competes has been crucial to the development of Silicon Valley's dynamic technology sector, facilitating the free flow of talent and ideas.[22] Studies have suggested that this policy has contributed to increased employee mobility, knowledge spillovers, and overall innovation in California's technology sector.[23] However, critics contend that this approach leaves businesses without adequate protection for legitimate interests and may discourage investment in employee development.[24]

2. The Reasonableness Approach

Most states have adopted a more nuanced "reasonableness" test for evaluating non-compete agreements. While the specific factors vary by jurisdiction, courts typically examine: (1) the scope of prohibited activities; (2) geographic limitations; (3) duration of restrictions; (4) the employee's access to confidential information or customer relationships; (5) the impact on the employee's ability to earn a living; and (6) the public interest.[25]

States differ significantly in how they apply these factors and in their willingness to modify or "blue pencil" overbroad agreements. Some jurisdictions, like New York, will modify unreasonable restrictions to make them enforceable, while others,

---

20 CAL. BUS. & PROF. CODE § 16600(a) (West 2025).

21 Edwards v. Arthur Andersen LLP, 189 P.3d 285, 290–91 (Cal. 2008).

22 For a discussion on the importance of competition in Silicon Valley, see ANNALEE SAXENIAN, REGIONAL ADVANTAGE: CULTURE AND COMPETITION IN SILICON VALLEY AND ROUTE 128, at 29–57 (1996).

23 Ronald J. Gilson, *The Legal Infrastructure of High Technology Industrial Districts: Silicon Valley, Route 128, and Covenants Not to Compete*, 74 N.Y.U. L. REV. 575, 575–76 (1999).

24 *See, e.g.*, Jonathan M. Barnett & Ted Sichelman, *The Case for Noncompetes*, 87 U. CHI. L. REV. 953, 953–54, 964–65 (2020).

25 RESTATEMENT (SECOND) OF CONTS. § 188 cmts. b–d (AM. L. INST. 1981).

like Wisconsin, will invalidate the entire agreement if any provision is unreasonable.[26]

3. Recent State Reform Efforts

In recent years, several states have enacted legislation to limit the use of non-compete agreements, particularly for low-wage workers. For example, Illinois prohibited non-competes for employees earning less than $75,000 annually (adjusted for inflation) through the Freedom to Work Act.[27] In 2018, Massachusetts enacted comprehensive non-compete reform, requiring garden leave payments[28] and limiting duration to twelve months.[29] Washington passed legislation in 2019 voiding non-competes for employees earning less than $100,000 annually and independent contractors earning less than $250,000.[30] These reforms reflect growing concern about the overuse of non-compete agreements and their potential to suppress wages and limit economic opportunity, particularly for lower-skilled workers.

C. The Federal Trade Commission's Rule

1. Overview and Legal Basis

The FTC's 2024 rule marks a dramatic shift from the traditional state-based approach to non-compete regulation. Citing its authority under section 5 of the FTC Act (Section 5) to prevent "unfair methods of competition," the FTC declared most non-compete agreements to be an unfair method of competition.[31] The rule prohibits employers from entering into or attempting to enter into non-compete agreements with workers, requires employers to rescind existing non-compete agreements, mandates notice to workers that existing non-compete agreements are no

---

26 *Compare* BDO Seidman v. Hirshberg, 712 N.E.2d 1220, 1226 (N.Y. 1999) ("[I]f the employer demonstrates an absence of overreaching, coercive use of dominant bargaining power, or other anti-competitive misconduct, but has in good faith sought to protect a legitimate business interest, consistent with reasonable standards of fair dealing, partial enforcement may be justified."), *with* Star Direct, Inc. v. Dal Pra, 767 N.W.2d 898, 916 (Wis. 2009) ("[I]f a restraint is unreasonable, the rest of that covenant is also unenforceable.").

27 820 ILL. COMP. STAT. 90/10(a) (2025).

28 A garden leave provision "requires an employee to remain employed while the employer is not obligated to assign work and restricts the employee from working for competitors." Garden Leave Provision, Practical Law Standard Clauses w-008-3138.

29 MASS. GEN. LAWS, ch. 149, § 24L(b)(iv), (b)(vii) (2024).

30 WASH. REV. CODE §§ 49.62.020(1)(b), 49.62.030(1) (2025).

31 FTC Non-Compete Clause Rule, 89 Fed. Reg. 38342, 38342, 38349 (May 7, 2024) (codified at 16 C.F.R. pt. 910) ("Alongside section 5, Congress adopted section 6(g) of the Act, in which it authorized the Commission to 'make rules and regulations for the purpose of carrying out the provisions of' the FTC Act, which include the Act's prohibition of unfair methods of competition.").

longer enforceable, and provides limited exceptions for non-competes related to the sale of businesses.[32]

2. Key Provisions and Implementation

The rule defines non-compete clauses broadly to include both explicit restrictions on competitive employment and de facto non-compete clauses that effectively preclude workers from seeking or accepting competitive employment.[33] However, it preserves employers' ability to protect trade secrets through other means, such as non-disclosure agreements and non-solicitation provisions, provided these do not function as de facto non-competes.[34]

Implementation requirements include (1) a 120-day compliance period for employers to rescind existing non-compete agreements, (2) mandatory worker notification requirements, (3) specific language requirements for rescission notices, and (4) record-keeping requirements for compliance documentation.[35]

3. Initial Market Response and Legal Challenges

The business community's response to the FTC rule has been mixed, with some industry groups promptly challenging its constitutionality.[36] Primary concerns include questions about the FTC's statutory authority to issue such a broad rule; constitutional challenges under the major questions doctrine; practical implementation difficulties, particularly for multi-state employers; and the potential impact on business investment and innovation.

As previously discussed, *Ryan, LLC v. FTC* resulted in a nationwide injunction and prompted the FTC to pause its appeal

---

32 FTC Non-Compete Clause Rule, 16 C.F.R. § 910.2 (2024).

33 *Id.* § 910.1 (defining non-compete clause).

34 Non-disclosure agreements are confidentiality agreements intended to protect against the unauthorized disclosure of trade secrets, ideas, and other confidential information to third parties. *See generally* Rex N. Alley, *Business Information and Nondisclosure Agreements: A Public Policy Framework*, 116 NW. U. L. REV. 817 (2021). Non-solicitation provisions restrict a party from soliciting employees, clients, or vendors and are assessed under the rule of reason to balance anti-competitive effects against pro-competitive justifications. THOMSON REUTERS, NON-SOLICITATION AND NO-POACH AGREEMENTS (2025), Westlaw 3-600-9465. Such provisions may be lawful if they are ancillary to a legitimate business interest. *See id.*

35 16 C.F.R. § 910.2(b); *see also id.* § 910.6 (establishing that the effective date of the FTC Non-Compete Clause Rule is September 4, 2024, which is 120 days after the rule was issued on May 7, 2024).

36 *U.S. Chamber to Sue FTC over Unlawful Power Grab on Noncompete Agreements Ban*, U.S. CHAMBER OF COM. (Apr. 23, 2024), https://www.uschamber.com/antitrust/u-s-chamber-to-sue-ftc-over-unlawful-power-grab-on-noncompete-agreements-ban [https://perma.cc/NZM6-U48J].

following a change in agency leadership.[37] But even before that, legal uncertainty had already taken hold. A federal court in Pennsylvania upheld the rule, concluding that the FTC acted within its statutory authority[38]—directly conflicting with the Texas decision and creating a circuit split. These divergent rulings reflect the fractured legal landscape surrounding the FTC's rule.

III. CONSTITUTIONAL AND FEDERAL AUTHORITY ISSUES

The transition from state-based regulation of non-compete agreements to federal oversight raises fundamental questions about the scope of federal authority and the balance between state and federal power. The FTC rule banning non-compete agreements represents an unprecedented federal intrusion into an area traditionally governed by state law, demanding careful constitutional analysis.[39] This analysis becomes particularly relevant as Congress considers legislative alternatives to the FTC's approach, including the potential adoption of a German-style compensatory system.

A. Commerce Clause Analysis

The constitutional foundation for federal regulation of non-compete agreements rests primarily on Congress's Commerce Clause authority.[40] Since *NLRB v. Jones & Laughlin Steel Corp.*, courts have interpreted this power broadly to include regulation of activities that substantially affect interstate commerce.[41] While traditionally governed by state law, non-compete agreements increasingly affect interstate commerce through their impact on national labor markets and innovation networks.[42]

Modern Commerce Clause jurisprudence suggests three potential bases for federal regulation of non-compete agreements. First, these agreements directly affect employee mobility across state lines, implicating Congress' authority to regulate channels of interstate commerce.[43] Second, non-compete restrictions influence

---

37 *See supra* Part I.

38 ATS Tree Servs., LLC v. FTC, No. 24-1743, slip op. at 18–19 (E.D. Pa. July 23, 2024).

39 *See* FTC Non-Compete Clause Rule, 89 Fed. Reg. 38342, 38355 (May 7, 2024) (codified at 16 C.F.R. pt. 910).

40 U.S. CONST. art. I, § 8, cl. 3.

41 NLRB v. Jones & Laughlin Steel Corp., 301 U.S. 1, 37 (1937).

42 *See* ALAN B. KRUEGER & ERIC A. POSNER, A PROPOSAL FOR PROTECTING LOW-INCOME WORKERS FROM MONOPSONY AND COLLUSION 13–15 (2018).

43 United States v. Lopez, 514 U.S. 549, 558 (1995) ("Congress may regulate the use of the channels of interstate commerce. . . . [and] Congress is empowered to regulate and protect the instrumentalities of interstate commerce, or persons or things in interstate commerce . . . .").

the national labor market, particularly in industries integral to interstate commerce, such as technology and financial services.[44] Third, and most significantly, the cumulative effect of non-compete agreements on labor markets, innovation, and economic competition provides a strong foundation for federal authority under the substantial effects doctrine.[45]

Under *Wickard v. Filburn*'s aggregation principle, Congress may regulate purely local activities if their cumulative effect substantially impacts interstate commerce.[46] Recent empirical studies estimate that twenty percent of American workers are bound by non-compete agreements, with significant effects on wage growth, job mobility, and economic competition.[47] This cumulative impact provides compelling justification for federal regulation, even when individual agreements might appear purely local in nature.

B. Constitutional Challenges to the Federal Trade Commission's Rule

The Supreme Court's recent emphasis on the major questions doctrine poses significant challenges to the FTC's authority to ban non-compete agreements.[48] Under this doctrine, courts expect Congress to "speak clearly if it wishes to assign to an agency decisions of vast economic and political significance."[49] The Court's decisions in *West Virginia v. EPA* and *NFIB v. OSHA* suggest growing skepticism toward broad agency interpretations of general statutory authority in economically significant matters.[50]

The FTC's reliance on Section 5's "unfair methods of competition" language to regulate employment contracts presents several vulnerabilities under the major questions doctrine.[51] First, the regulation of non-compete agreements represents an issue of vast economic significance, affecting millions of workers and

---

44 *See* Evan P. Starr, J.J. Prescott & Norman D. Bishara, *Noncompete Agreements in the U.S. Labor Force*, 64 J.L. & ECON. 53, 55–57 (2021).

45 *See* KRUEGER & POSNER, *supra* note 42, at 14.

46 Wickard v. Filburn, 317 U.S. 111, 128–29 (1942).

47 U.S. DEP'T OF THE TREASURY, THE STATE OF LABOR MARKET COMPETITION 28–29 (2022) ("[A] recent paper estimates that one-in-five workers is currently subject to non-compete agreements and double that number report having been bound by a non-compete agreement in the past.").

48 West Virginia v. Env't Prot. Agency, 597 U.S. 697, 723–24 (2022).

49 *Id.* at 716.

50 *See* NFIB v. OSHA, 595 U.S. 109, 117 (2022) (per curiam).

51 *See* FTC Non-Compete Clause Rule, 89 Fed. Reg. 38342, 38342 (May 7, 2024) (codified at 16 C.F.R. pt. 910) (citing 15 U.S.C. § 45(a)(1)).

thousands of businesses across all sectors of the economy.[52] Second, the FTC's assertion of authority over employment contracts marks a significant expansion of its traditional antitrust jurisdiction.[53] Third, the regulation of employment relationships has historically been left to the states or to specific congressional action, making this a matter of deep political significance.[54]

Beyond major questions concerns, the non-delegation doctrine presents additional constitutional challenges to the FTC's rule.[55] Although the Supreme Court has rarely invalidated congressional delegations of authority, the broad scope of the non-compete ban, coupled with limited statutory guidance in Section 5, raises questions about whether Congress provided an "intelligible principle" to guide the FTC's rulemaking.[56] Justice Gorsuch's dissent in *Gundy v. United States* signals renewed judicial interest in non-delegation constraints, particularly where agencies claim broad economic authority under general statutory provisions.[57]

The FTC's rule also faces substantial administrative law challenges under the Administrative Procedure Act.[58] Courts reviewing the rule will likely scrutinize whether the FTC adequately considered alternative regulatory approaches, state-level experimentation, industry-specific impacts, and compliance costs.[59] The D.C. Circuit's recent skepticism toward agency assertions of novel authority suggests particular attention to the FTC's statutory interpretation and cost-benefit analysis.[60]

C. Congressional Authority and Preemption

Unlike administrative agencies, Congress retains broader authority to regulate non-compete agreements, provided it acts within Commerce Clause boundaries. Congressional action could take several forms, ranging from complete preemption following California's model to establishing federal baseline requirements

---

52 *Id.* at 38344.

53 *Id.* at 38355–56.

54 *Id.* at 38452.

55 *See* Gundy v. United States, 588 U.S. 128, 166–67 (2019) (Gorsuch, J., dissenting).

56 *See* Gary S. Lawson, *"I'm Leavin' It (All) Up to You": Gundy and the (Sort-of) Resurrection of the Subdelegation Doctrine*, 2019 CATO SUP. CT. REV. 31, 33–35.

57 *See Gundy*, 588 U.S. at 149–69.

58 *See* 5 U.S.C. § 706(2)(A).

59 *See* Alden Abbott & Liya Palagashvili, *Policy Spotlight: The Problem with a Federal Ban on Noncompete Agreements*, MERCATUS CTR. (May 23, 2023), https://www.mercatus.org/research/policy-briefs/policy-spotlight-problem-federal-ban-noncompete-agreements [https://perma.cc/5N3R-6DQS].

60 *See* Am. Lung Ass'n v. Env't Prot. Agency, 985 F.3d 914, 930–32, 995 (D.C. Cir. 2021).

while allowing state variation.[61] The German compensatory model, if adopted through federal legislation, would likely face fewer constitutional hurdles than the FTC's outright ban.

Any federal legislation must carefully navigate preemption issues to avoid disrupting established state employment law frameworks.[62] The Supreme Court's preemption jurisprudence suggests several potential approaches to federal non-compete regulation.[63] Congress could explicitly preempt state non-compete law, though political considerations and federalism concerns might favor a more nuanced approach.[64] Alternatively, Congress might establish minimum standards while allowing states to impose additional restrictions, following the model of other federal employment laws, such as the Fair Labor Standards Act (FLSA).[65]

The FLSA mandates that employers must pay minimum wage and overtime compensation; it includes provisions outlining remedies, as well as various exemptions from its coverage.[66] Notably, the FLSA contains a saving clause that permits states to adopt and enforce minimum wage standards that are more stringent than those set by federal law.[67]

The German compensatory model presents particularly complex preemption challenges. Implementation would require careful consideration of how federal compensation requirements interact with existing state restrictions on non-compete agreements.[68] For example, states like California might maintain their outright prohibitions as more protective of employee rights, while other states could adopt the federal compensation framework as a baseline.[69]

D. Federalism Concerns

The regulation of non-compete agreements implicates core federalism principles, particularly given states' traditional

---

61 *See* Barnett & Sichelman, *supra* note 24, at 958–62.

62 *See* Roderick M. Hills, *Against Preemption: How Federalism Can Improve the National Legislative Process*, 82 N.Y.U. L. REV. 1, 4–6 (2007).

63 *See* Catherine M. Sharkey, *Inside Agency Preemption*, 110 MICH. L. REV. 521, 524–26 (2012).

64 *See* Altria Group, Inc. v. Good, 555 U.S. 70, 76–77 (2008).

65 *See* Daniel V. Dorris, *Fair Labor Standards Act Preemption of State Wage-and-Hour Law Claims*, 76 U. CHI. L. REV. 1251, 1258 (2009).

66 Fair Labor Standards Act, 29 U.S.C. §§ 201–219.

67 *Id.* § 218.

68 *See* Katherine V.W. Stone, *The New Psychological Contract: Implications of the Changing Workplace for Labor and Employment Law*, 48 UCLA L. REV. 519, 645–51 (2001).

69 *See* Gilson, *supra* note 23, at 578–80.

authority over employment relationships.[70] The Supreme Court has repeatedly emphasized that federal intrusion into areas of traditional state regulation requires careful justification, especially where state-level experimentation has produced varied regulatory approaches.[71] This "laboratory of democracy" function has particular salience in the non-compete context, where states have developed diverse approaches reflecting local economic conditions and policy preferences.

Implementation of a German-style system would require careful attention to state-federal coordination. Unlike Germany's unitary system, the American federal structure necessitates mechanisms for coordinating between federal standards and state enforcement. This coordination becomes particularly important in areas, such as anti-commandeering and state innovation.

Enforcement mechanisms must respect anti-commandeering principles while ensuring effective implementation.[72] Congress cannot compel state officials to enforce federal non-compete regulations, but it can provide incentives for state participation and establish concurrent enforcement authority.[73] The experience of other federal employment laws suggests that a cooperative enforcement model, with clear division of authority between federal and state agencies, offers the most promising approach.[74]

Additionally, any federal framework must preserve appropriate spheres for state innovation. While establishing uniform compensation requirements, federal legislation should maintain state authority to address industry-specific concerns and local economic conditions.[75] This flexibility proves particularly important given the significant variations in regional labor markets and industry concentrations across states.

E. Constitutional Framework for a German-Style System

The implementation of a German-style compensatory system requires careful constitutional structuring to ensure both effectiveness and legitimacy within the American federal system. Unlike the FTC's categorical ban, a compensatory approach

---

70 *See* Gregory v. Ashcroft, 501 U.S. 452, 458 (1991).

71 *See* New State Ice Co. v. Liebmann, 285 U.S. 262, 311 (1932) (Brandeis, J., dissenting).

72 Murphy v. Nat'l Collegiate Athletic Ass'n, 584 U.S. 453, 469–73 (2018).

73 *See* Jessica Bulman-Pozen & Heather K. Gerken, *Uncooperative Federalism*, 118 YALE L.J. 1256, 1258–60 (2009).

74 *See* Benjamin I. Sachs, *Despite Preemption: Making Labor Law in Cities and States*, 124 HARV. L. REV. 1153, 1157–58 (2011).

75 *See* David A. Super, *Rethinking Fiscal Federalism*, 118 HARV. L. REV. 2544, 2547–49 (2005).

better aligns with American constitutional traditions of protecting both economic liberty and property rights while regulating market excesses.

Congress' authority to implement such a system would likely derive from three constitutional sources. First, the Commerce Clause provides primary authority given the direct impact of non-compete agreements on interstate labor markets and economic competition. Second, the Necessary and Proper Clause offers additional support for administrative mechanisms needed to implement the compensation system.[76] Third, Congress' power to establish inferior federal courts enables creation of specialized adjudicative procedures for non-compete disputes.[77]

The constitutional architecture of a German-style system should include the following several key elements. First, clear jurisdictional boundaries must define the scope of federal oversight. The system should apply to non-compete agreements that either directly involve interstate commerce, affect employees in industries substantially connected to interstate commerce, or have cumulative effects on national labor markets.[78] This jurisdictional framework would satisfy Commerce Clause requirements while respecting traditional state authority over purely local employment matters.[79] Second, the system must establish appropriate administrative structures while avoiding non-delegation concerns. Unlike the FTC's broad rulemaking approach, legislation implementing a German-style system should provide detailed statutory standards for minimum compensation requirements, duration limitations, industry-specific adjustments, enforcement mechanisms, and judicial review standards. Third, the framework must incorporate adequate procedural protections to satisfy due process requirements. These protections should include notice requirements for affected employees, hearing rights for compensation disputes, appeal mechanisms for administrative determinations, and judicial review of enforcement actions. Finally, any federal framework must preserve state constitutional interests while achieving national

76 M'Culloch v. Maryland, 17 U.S. (4 Wheat.) 316, 324–25 (1819).

77 *See* Daniel J. Meltzer, *Legislative Courts, Legislative Power, and the Constitution*, 65 IND. L.J. 291, 292 (1990).

78 *See* Gonzales v. Raich, 545 U.S. 1, 17–19 (2005).

79 *See generally* Larry Kramer, *Understanding Federalism*, 47 VAND. L. REV. 1485, 1488–90 (1994).

uniformity.[80] This balance requires careful attention to several federalism principles.

The system should maintain state courts' authority to adjudicate non-compete disputes while ensuring uniform application of federal standards.[81] States should retain power to enforce more protective standards, following the model of other federal employment laws. Additionally, the framework should provide a foundation for implementing German-style reforms while respecting American constitutional traditions and federal structure. Unlike the FTC's categorical approach, a carefully structured compensatory system offers a path forward that better aligns with both constitutional requirements and practical realities of American labor markets.[82]

IV. COMPARATIVE ANALYSIS OF ALTERNATE MODELS AND THE GERMAN MODEL

A. The United Kingdom's Approach

The United Kingdom (UK) does not outright ban non-compete agreements but strictly regulates them to ensure proportionality. Under UK law, non-competes are enforceable only if they serve a legitimate business interest, such as protecting trade secrets or preventing unfair competition. However, UK courts assess them under the principle of reasonableness, evaluating the duration, geographic scope, and industry restrictions.[83] Employers often use garden leave, where employees remain on the payroll during the restricted period but are prevented from working for competitors.[84]

The UK model demonstrates that rigorous judicial oversight and limited enforceability of non-competes can strike a balance between business protection and employee mobility. While this approach prevents outright abuse, it still places the burden on employees to challenge agreements, potentially deterring low- to mid-level workers from contesting unfair restrictions.

---

80 *See* Gillian E. Metzger, *Administrative Law as the New Federalism*, 57 DUKE L.J. 2023, 2026–28 (2008).

81 *Id.*

82 *See* Cynthia Estlund, *Labor Law Reform Again? Reframing Labor Law as a Regulatory Project*, 16 N.Y.U. J. LEGIS. & PUB. POL'Y 383, 386–87 (2013).

83 *See* Catriona Watt & Maree Cassaidy, *The UK Perspective on the FTC Ban on US Non-Competes*, INT'L BAR ASS'N (Aug. 27, 2024), https://www.ibanet.org/UK-perspective-on-the-FTC-ban-on-US-non-competes [https://perma.cc/4XQ7-RRTT].

84 *See id.*

B. The Canadian Model

Canada follows a similar reasonableness approach, with significant variation across provinces. In Ontario, non-competes were largely prohibited under the 2021 Working for Workers Act, except in cases of senior executives.[85] Other provinces, such as British Columbia and Alberta, have taken a cautious approach, requiring non-competes to be clear, limited in scope, and demonstrably necessary.[86]

Canadian courts have also shown strong preference for alternative restrictive covenants, such as non-solicitation and confidentiality agreements, over outright non-competes.[87] The Canadian approach suggests that a nationwide standard, with limited exceptions, can enhance worker mobility while protecting business interests.

C. Alternative U.S. Proposals: Workforce Mobility Act and State-Level Innovations

The Workforce Mobility Act (WMA), a federal proposal introduced in the U.S. Senate, aims to prohibit most non-competes except in limited cases, such as the sale of a business.[88] This legislative effort aligns with California's complete ban and the FTC's rule, arguing that non-competes suppress wages and hinder innovation. While the WMA addresses concerns about the restrictive nature of non-compete agreements, it fails to provide the nuanced balance needed to protect both worker mobility and legitimate business interests. The WMA's near-total prohibition on non-competes does not accommodate industries that rely on these agreements to protect proprietary information, customer relationships, and long-term investments in employee training.[89]

---

85 Working for Workers Act, 2021, S.O. 2021, P.35 (Can. Ont.) (discussing prohibitions on non-competes but providing an exception for executives).

86 The enforceability of non-compete clauses in British Columbia and Alberta is determined under the common law. *See Are Non-Compete Clauses Enforceable in British Columbia?*, TAYLOR JANIS WORKPLACE L., https://www.tjworkplacelaw.com/blog/bc/are-non-compete-clauses-enforceable-in-bc/ [https://perma.cc/E2N2-Q3PE] (Sept. 26, 2024); David Di Gianvittorio, *Non-Competition Covenants in Alberta*, FIELD L. (May 2024), https://www.fieldlaw.com/News-Views-Events/232902/Non-Competition-Covenants-in-Alberta [https://perma.cc/27HZ-9H78].

87 *See* Anita Nador & Jordan Epstein, *U.S. Non-Compete Agreements Ban: Parallels in Canada*, HR.COM (Oct. 3, 2024), https://ide.hr.com/en/magazines/all_articles/us-non-compete-agreements-ban-parallels-in-canada_m1sxilzx.html?s=22ynpGBmd1T1r7CV9Zj [https://perma.cc/L7ZM-R4GX].

88 Workforce Mobility Act of 2023, S. 220, 118th Cong. (2023).

89 *See* Starr, Prescott & Bishara, *supra* note 44, at 53–55.

Moreover, the WMA does not solve the patchwork problem currently plaguing state-level enforcement.[90] While it preempts state laws, its one-size-fits-all approach fails to account for the fact that some states have already implemented reasonable restrictions rather than outright bans.[91] States like Massachusetts and Illinois have demonstrated that compromise solutions—such as requiring compensation for non-competes—can be effective in balancing interests.[92]

Germany's sophisticated approach offers valuable insights for U.S. reform efforts.[93] The German model, developed over more than a century, represents one of the most balanced and effective regulatory frameworks globally for managing post-employment competition restrictions.[94]

D. Legal Framework and Historical Development of the German Model

The German approach to post-employment non-compete agreements—*nachvertragliche Wettbewerbsverbote*—reflects the country's broader commitment to social market economics and cooperative labor relations.[95] Codified in Sections 74–75f of the German Commercial Code, *Handelsgesetzbuch* (HGB), this system has evolved through legislative refinement and judicial interpretation to create a careful equilibrium between employer and employee interests.[96]

The modern framework emerged from nineteenth-century commercial law reforms, particularly the 1897 Commercial Code, which first established basic principles for regulating trade restrictions. Following World War II, the newly established Federal Republic integrated non-compete regulation into its distinctive social market economy (*Soziale Marktwirtschaft*),

---

90 *See State Noncompete Law Tracker*, ECON. INNOVATION GRP. (Oct. 11, 2024), https://eig.org/state-noncompete-map [https://perma.cc/W6GG-YJA8].

91 *Id.*

92 *See, e.g.*, MASS. FED. LAWS ch. 149, § 24L (2018); 820 ILL. COMP. STAT. 90/5–15 (2022).

93 *See* JENS KIRCHNER ET AL., KEY ASPECTS OF GERMAN EMPLOYMENT AND LABOUR LAW 161–70 (2d ed. 2018).

94 *See* 2 WOLFGANG DÄUBLER, DAS ARBEIRSRECHT 1023–25 (12th ed. 2019); *see also* Michael Schley, *Restrictive Covenants – Key Considerations for Employees in Germany – November 2022*, DWF (Oct. 11, 2022), https://dwfgroup.com/en/news-and-insights/insights/2022/10/restrictive-covenants-key-considerations-for-employers-in-germany [https://perma.cc/C7Q9-9CAG].

95 *See* Manfred Weiss, *The Interface Between Constitution and Labor Law in Germany*, 26 COMPAR. LAB. L. & POL'Y J. 181, 192–96 (2005).

96 Handelsgesetzbuch [HGB] [Commercial Code], §§ 74–75f, https://www.gesetze-im-internet.de/englisch_hgb/englisch_hgb.html [https://perma.cc/3TDZ-6686] (Ger.).

emphasizing both economic freedom and social protection.[97] This integration reflected Germany's constitutional commitment to occupational freedom (*Berufsfreiheit*) under Article 12 of the Basic Law (*Grundgesetz*), while recognizing legitimate business interests in protecting confidential information and customer relationships.[98]

Four fundamental principles distinguish the German approach. First, employers must provide mandatory compensation (*Karenzentschädigung*) of at least fifty percent of the employee's total remuneration during the restriction period.[99] Second, non-compete agreements cannot exceed two years in duration.[100] Third, restrictions must protect legitimate business interests and be proportionate in scope. Fourth, all agreements must be in writing and provide clear notice of restrictions and compensation.[101]

E. Core Requirements and Implementation

1. Mandatory Compensation Framework

The cornerstone of the German system is its mandatory compensation requirement, which transforms non-compete agreements from unilateral restrictions into bilateral economic arrangements. Under HGB section 74(2), employers must pay at least fifty percent of the employee's total remuneration during the restriction period,[102] a requirement that fundamentally alters the incentive structure for both parties.

The German Federal Labor Court has developed sophisticated standards for calculating this compensation. Total compensation encompasses base salary and all regular compensation components: bonuses, commissions, benefits, and other monetary advantages.[103] German courts have expanded this definition to

---

97 *See* Martin Behrens & Wade Jacoby, *The Rise of Experimentalism in German Collective Bargaining*, 42 BRIT. J. INDUS. REL. 95, 97–99 (2004); *see also* Gunther Schnabl, *The 1948 German Currency and Economic Reform: Lessons for European Monetary Policy*, 39 CATO J. 607, 608–14 (2019).

98 Grundgesetz [GG] [Basic Law] art. 12, translation at http://www.gesetze-im-internet.de/englisch_gg/ [https://perma.cc/KL7Y-RJZM] (Ger.).

99 Handelsgesetzbuch [HGB] [Commercial Code], § 74, para. 2, https://www.gesetze-im-internet.de/englisch_hgb/ [https://perma.cc/RSK9-8DAU] (Ger.).

100 Handelsgesetzbuch [HGB] [Commercial Code], §§ 74–74a, https://www.gesetze-im-internet.de/englisch_hgb/englisch_hgb.html [https://perma.cc/3TDZ-6686] (Ger.).

101 *Id.*

102 *Id.* § 74, para. 2.

103 Thilo Mahnhold, *Choice of Law Provisions in Contractual Covenants Not to Compete: The German Approach*, 31 COMPAR. LAB. L. & POL'Y J. 331, 332 (2010). For more details, see WETTBEWERBSVERBOTE (Jobst-Hubertus Bauer & Martin Diller eds., 4th ed. 2006).

include regular profit-sharing payments and long-term incentives, reflecting the court's emphasis on maintaining the employee's economic position.[104]

This compensation requirement serves multiple functions. First, it forces employers to internalize the cost of restrictions, leading to more efficient use of non-compete agreements. Second, it provides employees with economic security during the restriction period, facilitating compliance and reducing litigation.

2. Temporal and Scope Limitations

German law imposes strict temporal limits on non-compete agreements while requiring careful tailoring of their scope. The two-year maximum duration, established in HGB section 74a paragraph 1, reflects legislative balancing between protecting legitimate business interests and ensuring worker mobility.[105]

Germany has established a detailed framework for assessing geographic and activity restrictions in post-contractual non-compete agreements to ensure they protect legitimate business interests without unduly limiting an employee's future employment opportunities.[106] Key aspects of this framework include:

1. Activity Restrictions: The scope of prohibited activities must be clearly defined and limited to what is necessary to protect the employer's legitimate interests. Overly broad restrictions that prevent the employee from engaging in any work in the industry are generally considered unenforceable.[107]
2. Geographical Scope: The territorial limitations of a non-compete clause must align with the employer's actual

---

104 Mahnhold, *supra* note 103, at 332–33.

105 Handelsgesetzbuch [HGB] [Commercial Code] § 74a, para. 1, https://www.gesetze-im-internet.de/englisch_hgb/ [https://perma.cc/RSK9-8DAU] (Ger.). It is important to note that HGB section 74 outlines the general requirements for post-contractual non-compete clauses, including the necessity for compensation to the employee. However, the enforceability of contractual penalties within these agreements is addressed in HGB section 75c.

106 *See* Christian Maron & Benedikt Groh, *10 Pitfalls when Entering into a Post-Contractual Non-Compete Covenant Under German Law*, TAYLORWESSING (Dec. 3, 2020), https://www.taylorwessing.com/de/insights-and-events/insights/2020/12/10-pitfalls-when-entering-into-a-post-contractual-non-compete-covenant-under-german-law [https://perma.cc/9Y8S-5X3Y].

107 *A Comparison of Laws in Selected EU Jurisdictions Relating to Post-Contractual, Non-Competition Agreements Between Employers and Employees*, NORTON ROSE FULBRIGHT (Aug. 2017), https://www.nortonrosefulbright.com/en/knowledge/publications/9807eea3/a-comparison-of-laws-in-selected-eu-jurisdictions-relating-to-post-contractual-non-competition-agreements-between-employers-and-employees#section3 [https://perma.cc/AKW9-UY28] ("The covenant

business operations. Restrictions extending beyond the regions where the employer operates are typically deemed unreasonable and thus unenforceable.[108]

3. Duration: German law caps the duration of post-contractual non-compete clauses at a maximum of two years. Any agreement exceeding this period is considered non-binding, allowing the employee to choose whether to adhere to the clause.[109]
4. Compensation: The compensation provided during the non-compete period (*Karenzentschädigung*) is processed through standard payroll procedures, with applicable deductions for taxes and social security contributions. It must be precisely calculated based on the employee's most recent salary, including any regular bonuses or other recurring compensatory elements that were part of their income.[110]

These regulations collectively ensure that non-compete clauses in Germany are balanced, safeguarding the employer's interests while preventing unreasonable constraints on the employee's professional future.

F. Enforcement Mechanisms and Judicial Oversight

1. Procedural Framework

The German system provides sophisticated enforcement mechanisms through specialized labor courts (*Arbeitsgerichte*) with expertise in employment matters.[111] These courts operate under expedited procedures designed to resolve disputes quickly while ensuring thorough consideration of both parties' interests.

The enforcement of labor disputes in Germany typically begins with a conciliation hearing (*Gütetermin*) before the labor

must be reasonable and can be included only to protect the legitimate interests of the company.").

108 *Id.*

109 Bredin Prat et al., *Post-Contractual Restrictions on Competition in Employment Relationships in the UK, France and Germany*, HENGELER MUELLER (Dec. 2016), https://www.hengeler.com/de/service/newsletter/december-2016 [https://perma.cc/987C-3CJU].

110 *Non-Compete Clauses in Employment Contracts: Legal and Payroll Considerations*, PAYROLL GERMANY, https://payrollgermany.de/blog/non-compete-clauses-in-employment-contracts-legal-and-payroll-considerations/ [https://perma.cc/3WFY-78Y6] (last visited Apr. 12, 2025).

111 *See Responsibilities of the Federal Labour Court*, BUNDESARBEITSGERICHT, https://www.bundesarbeitsgericht.de/responsibilities/ [https://perma.cc/6SUH-EAYZ] (last visited Apr. 12, 2025).

court's chamber, led by a judge.[112] This initial step aims to resolve disputes early and is a core feature of the German labor process. While not formally classified as mandatory mediation, these hearings often result in settlement, contributing to reduced litigation burdens. The Federal Labor Court's 2021 report confirms the importance of this procedural design, though it does not cite specific resolution rates.[113]

If conciliation is unsuccessful, the dispute proceeds to a full hearing before the labor court's panel.[114] While some cases are resolved promptly, complex matters—particularly those involving appeals—may extend beyond six months, and in rare cases, over two years.[115] Appeals are handled by specialized Higher Labor Courts (*Landesarbeitsgerichte*) and, if necessary, by the Federal Labor Court (*Bundesarbeitsgericht*), which ensures consistency in the interpretation of employment law standards, including post-contractual non-compete provisions.[116]

2. Remedies and Enforcement

The German system provides a comprehensive framework of remedies designed to ensure effective enforcement while promoting compliance. Unlike the American approach, which often relies primarily on injunctive relief, German law offers a more nuanced array of remedial options.

a. Monetary Remedies

Courts can award several types of monetary relief. When an employee breaches a valid post-contractual non-compete agreement, the employer has several monetary remedies available. The employer can immediately stop paying the agreed compensation (*Karenzentschädigung*) for the duration of the breach.[117] The employer may also claim damages from the

---

112 *See What Is a Conciliation Hearing (Gütermin)?*, FAIRE INTEGRATION, https://www.faire-integration.de/en/article/1041.how-can-i-get-a-consulting-certificate.html [https://perma.cc/8MJ5-P3G8] (last visited May 18, 2025).

113 *See* Bundesarbeitsgericht [BAG] [Federal Labor Court], Jahresbericht 1, 45–47 (2021) (Ger.).

114 *See What Is a Conciliation Hearing (Gütermin)?*, *supra* note 112.

115 *Id.*

116 *Id.*; *see also Responsibilities of the Federal Labour Court*, *supra* note 111.

117 *Restrictive Covenants — Key Considerations for Employers in Germany — November 2022*, DWF GRP. (Oct. 11, 2022), https://dwfgroup.com/en/news-and-insights/insights/2022/10/restrictive-covenants-key-considerations-for-employers-in-germany [https://perma.cc/6A4Y-9BB7]; *see also Compensation for Non-Compete/Karenzentschädigung*, KUHLEN (May 25, 2022), https://www.kuhlen-berlin.de/en/glossary/karenzentschaedigung [https://perma.cc/4EUA-GEHZ].

employee, which can include compensation for lost profits resulting from the breach.[118] If the non-compete agreement includes a penalty clause for breaches, German labor courts will enforce it, obligating the employee to pay the stipulated penalty.[119] These remedies aim to protect the employer's legitimate business interests and deter employees from violating their contractual obligations.

b. Injunctive Relief

Germany approaches injunctive relief with greater flexibility than its American counterparts.[120] Several forms of injunctive relief are available. Regarding preventive and prohibitive measures, employers can seek court orders to prevent a former employee from initiating competitive activities or to prohibit the continuation of such activities if they have already commenced. This proactive approach helps mitigate potential harm to the employer's business.[121]

The employer also has procedural mechanisms at their disposal. Given the time-sensitive nature of competitive breaches, employers may file for preliminary injunctions (*einstweilige Verfügung*) to obtain court orders restraining the former employee's competitive actions.[122] This expedited process addresses the urgency of the situation. Alternatively, employers may initiate standard legal proceedings to secure a permanent injunction against the former employee's competitive conduct.

Relief is available so long as there exists a valid non-compete agreement that complies with German legal standards (including appropriate compensation and reasonable limitations concerning duration, geographic scope, and activity restrictions),[123] an imminent or ongoing breach (meaning the employer must

---

118 *Restrictive Covenants — Key Considerations for Employers in Germany — November 2022*, *supra* note 117.

119 Handelsgesetzbuch [HGB] [Commercial Code], § 75c, https://www.gesetze-im-internet.de/englisch_hgb/ [https://perma.cc/RSK9-8DAU] (Ger.).

120 Britta Grauke, Gero Pogrzeba & Daniel Matijevic, *Injunctive Relief in German Civil Cases — No Longer Possible Without an Oral Hearing?*, EUR. DISPS. BLOG (Sept. 13, 2024), https://european-disputes-blog.weil.com/germany/injunctive-relief-in-german-civil-cases-no-longer-possible-without-an-oral-hearing/ [https://perma.cc/WVF6-GD2D].

121 Prat et al., *supra* note 109.

122 *See* Olaf Lampke, *§ 55 Einstweiliger Rechtsschutz / 8. Wettbewerbsverbot*, HAUFE, https://www.haufe.de/id/beitrag/55-einstweiliger-rechtsschutz-8-wettbewerbsverbot-HI15824104.html [https://perma.cc/TM4F-Z6JD] (last visited May 8, 2025) ("Der Anspruch des Arbeitgebers, dass der Arbeitnehmer Wettbewerbshandlungen unterlässt, kann im Wege der einstweiligen Verfügung geltend gemacht warden," which means: "The employer's claim that the employee refrains from competitive acts can be asserted by way of a preliminary injunction.").

123 Maron & Groh, *supra* note 106.

demonstrate that the former employee is either currently violating the non-compete agreement or is likely to do so imminently),[124] and the employer must continue to fulfill any obligations outlined in the non-compete agreement while pursuing injunctive relief.[125]

3. Oversight

The German system provides multiple layers of institutional oversight, distinguishing it from more court-centric American approaches. This oversight operates through several mechanisms.

First, works councils (*Betriebsräte*) play a crucial role in monitoring non-compete practices.[126] While works councils do not have direct oversight over individual non-compete agreements, they play an indirect role in monitoring broader employment practices related to fairness and employee protections. Under the Works Constitution Act, they have consultation rights on standard employment terms that might include non-compete policies applied across the workforce. However, enforcement and review of specific post-contractual non-compete agreements remain the domain of the courts and the individual parties involved.[127]

Second, specialized labor courts provide ongoing supervisory functions beyond dispute resolution.[128] These courts issue advisory opinions on proposed restrictions, monitor compliance with prior orders, review modification requests, coordinate with works councils, and maintain databases of precedential decisions.[129] Through specialized labor courts and structured supervisory functions, the German enforcement mechanisms and judicial oversight encourage early dispute resolution while ensuring balanced considerations of each party's interests and rights. This approach reduces the burden on the judiciary and advances key policy objectives of protecting employees, promoting legal certainty, and fostering equitable labor practices across sectors.

---

124 *See* Theresa Richter, *Contractual and Post-Contractual Non-Compete Obligations — When Does an Employee's Obligation to Refrain from Competition End?*, DLA PIPER (Apr. 25, 2024), https://www.dlapiper.com/en/insights/blogs/employment-blog-germany/2024/vertragliches-und-nachvertragliches-wettbewerbsverbot [https://perma.cc/M356-RXUL].

125 *Cf. Restrictive Covenants in Germany*, L&E GLOB. (Oct. 22, 2024), https://leglobal.law/countries/germany/employment-law/employment-law-overview-germany/08-restrictive-covenants/ [https://perma.cc/D3G6-XFS7].

126 Betriebsverfassungsgesetz [BetrVG] [Works Constitution Act] § 87, para. 1, https://www.gesetze-im-internet.de/englisch_betrvg/englisch_betrvg.html [https://perma.cc/Z6E8-SD77] (Ger.).

127 *Id.*

128 *See Responsibilities of the Federal Labour Court*, *supra* note 111.

129 *Id.*

Finally, Germany's model demonstrates that procedural expediency can be harmonized with substantive justice and institutional accountability.

G. Economic Impacts and Empirical Evidence

1. Market Effects and Innovation Outcomes

Germany's balanced framework has been associated with positive economic outcomes, particularly in terms of market efficiency and innovation. The approach serves as a driving force for competitive markets, entrepreneurial activity, and long-term productivity that preserve labor market efficiency by strict adherence to limitations on scope, duration, and compensation during non-compete periods.

In terms of market efficiency, the German system promotes both worker mobility and healthy competitive practices. First, the German system mandates that non-competes must be reasonable in scope and duration, with a maximum enforceable period of two years and a requirement for employers to provide at least fifty percent of the employee's last remuneration as compensation during the non-compete period.[130] These stringent requirements deter employers from imposing overly restrictive non-competes, thereby promoting higher worker mobility. Increased mobility allows employees to seek opportunities that better match their skills, leading to a more efficient allocation of labor across the market.[131] Second, by limiting the enforceability of overly restrictive non-competes, Germany fosters a competitive labor market where companies must continually innovate and improve working conditions to attract and retain talent. This competition among employers can lead to better job matches and enhanced productivity.

Germany's approach has also yielded strong innovation outcomes. One key factor is the facilitation of knowledge spillover through the movement of skilled professionals between firms. This mobility encourages the dissemination of knowledge and best practices across the industry, fostering an environment conducive to innovation. When employees transition between companies, they bring diverse experiences and ideas that can lead to novel solutions and advancements. Additionally, Germany's legal framework support entrepreneurial activity by ensuring that agreements are

130 Handelsgesetzbuch [HGB] [Commercial Code], § 74–74a, para. 1, https://www.gesetze-im-internet.de/englisch_hgb/ [https://perma.cc/RSK9-8DAU] (Ger.).

131 *Cf.* Köckeritz & Zeppenfeld, *supra* note 14.

not overly restrictive, Germany supports entrepreneurial endeavors. Employees with innovative ideas are more likely to establish startups without the fear of legal repercussions from former employers. This entrepreneurial activity contributes to a dynamic economy with continuous technological advancements.

Empirical research, while still developing, suggests that the German approach produces measurable economic benefits. Although, direct empirical studies on the impact of Germany's regulations are limited, broader research supports that stringent agreements can suppress wages and stifle innovation. Conversely, balanced approaches, such as Germany's, that impose reasonableness and compensation requirements, are associated with positive economic outcomes. These include higher job switching rates, greater startup formation, and improved productivity in knowledge-intensive sectors. As such, the empirical evidence reinforces the broader policy rationale behind Germany's compensatory framework.[132]

2. Labor Market Dynamics and Human Capital

Rather than simply restricting job transitions, the German system appears to *channel* mobility in ways that preserve labor market efficiency. Compensated non-competes may reduce the urgency to take suboptimal jobs, allowing employees to seek roles better aligned with their skills and aspirations. While direct empirical studies on labor mobility's impact are limited, broader research indicates that worker transitions often lead to improvements in job quality and wages, especially in knowledge-intensive sectors.[133] Additionally, scholars suggest that mandatory compensation during the non-compete period provides financial stability, enabling affected employees to invest in

132 *See Non-Compete Clauses in the UK and U.S.: Recent Trends*, COVINGTON (Sept. 2024), https://www.cov.com/en/news-and-insights/insights/2024/09/non-compete-clauses-in-the-uk-and-us-recent-trends [https://perma.cc/LH4M-CBVD].

133 *See* Anders Akerman & Kerstin Holzheu, *The Role of Workers in Knowledge Diffusion Across Firms* (Scis. Po Dep't Econ., Discussion Paper No. 2024-04, 2024), https://www.sciencespo.fr/department-economics/sites/sciencespo.fr.department-economics/files/2024_v2_akerman_holzheu_the_role_of_workers_in_knowledge_diffusion_across_firms.pdf [https://perma.cc/QUB4-W23S].

professional development, including continuing education and credentialing.[134]

German employers operating under enforceable non-competes have added incentive to invest in workforce development, since they can better retain and protect that investment. Studies show that enforceable non-competes reduce the risk of losing trained employees to competitors, which in turn encourages more investment in employee training and structured onboarding programs.[135] Also, German firms are more likely to use knowledge management systems, including formal documentation processes and mentorship frameworks, particularly in sectors where employee turnover is costly.[136]

H. Comparative Advantages and Implementation Challenges

The adoption of a structured and statutory framework for enforcing post-contractual non-compete agreements presents clear advantages over the current U.S. approach. By replacing ad hoc, discretionary enforcement with clear legal standards and defined penalties for non-compliance, such a framework would reduce litigation burdens and promote more consistent, predictable dispute resolution.[137]

A specialized adjudication process, as exemplified by the German labor court system, fosters greater compliance and deters misuse of restrictive covenants.[138] While implementation in the United States would face challenges, including fragmented

---

134 *See* Yann Richard & David Al Mari, *Should Non-Compete Clauses Be Compensated?*, ASS'N OF CORP. COUNS. (Apr. 24, 2014), https://www.acc.com/resource-library/should-non-compete-clauses-be-compensated [https://perma.cc/846N-G6BQ].

135 Starr, Prescott & Bishara, *supra* note 44, at 80–81.

136 UWE CANTNER, KRISTIN JOEL & TOBIAS SCHMIDT, THE EFFECTS OF KNOWLEDGE MANAGEMENT ON INNOVATIVE SUCCESS — AN EMPIRICAL ANALYSIS OF GERMAN FIRMS 1, 15–16 (2009), https://www.bundesbank.de/resource/blob/703494/3cc7421c7b9c09a9a80e65f71644b390/mL/2009-07-13-dkp-16-data.pdf [https://perma.cc/XHC9-GP5T].

137 *See* Kai Bodenstedt & Henriette Norda, *Germany's Post-Contractual Non-Compete Covenants in a Nutshell*, DLA PIPER, https://www.dlapiperaccelerate.com/knowledge/2017/germanys-post-contractual-non-compete-covenants-in-a-nutshell.html [https://perma.cc/C74S-4E3F] (last visited May 8, 2025) (describing statutory clarity as reducing litigation risk in Germany). In contrast, the prevailing U.S. model—featuring case-by-case judicial discretion in general civil courts—often produces fragmented outcomes and legal uncertainty. *See* Starr, Prescott & Bishara, *supra* note 44, at 68, 69 fig.8 (illustrating the variation in judicial enforcement of non-competes across U.S. jurisdictions).

138 *See* Peter Hanau, *Das Bundesarbeitsgericht*, BUNDESARBEITSGERICHT, https://www.bundesarbeitsgericht.de/ [https://perma.cc/6N9V-AH6E] (last visited May 8, 2025) (detailing the role of specialized labor courts in early conciliation and resolution of employment disputes).

state-level authority, limited judicial specialization, and the costs of enhanced oversight, these obstacles are not insurmountable.[139] A phased, factor-based transition strategy could address institutional concerns while allowing time for adaptation.

Employers would need to adjust to new financial obligations—such as compensating employees during the restricted period[140]—but the long-term benefits in legal clarity, administrative efficiency, and labor mobility strongly favor reform. Overall, this model offers a compelling path forward for modernizing non-compete regulation in a way that balances economic flexibility with worker protections.

1. Systemic Advantages Over Alternative Approaches

Germany's regulatory framework offers distinct advantages over both categorical prohibitions and traditional common law approaches. These advantages manifest in areas such as economic efficiency, enforcement costs, and stakeholder outcomes.

a. Reduced Litigation and Efficient Dispute Resolution

With Germany's clear statutory requirements for non-competes—the mandatory compensation requirement and reasonableness criteria—it fosters a more predictable and efficient legal environment. The compensation requirement in particular creates more balanced incentives compared to jurisdictions that use binary prohibitions or subjective reasonableness tests. Because employers are required to pay for the restriction, they are compelled to assess the true economic value of each non-compete. This leads to more selective and intentional use of such agreements.[141] The framework would also introduce a new cost burden associated with non-competes, which might well have the effect of discouraging blanket non-compete clauses and fosters more thoughtful deployment of restrictive covenants, which improves fairness and economic efficiency.[142] In addition, clear statutory guidelines, coupled with predictable

139 *See* U.S. DEP'T OF THE TREASURY, NON-COMPETE CONTRACTS: ECONOMIC EFFECTS AND POLICY IMPLICATIONS 15–16 (2016).

140 *Handelsgesetzbuch* [HGB] [Commercial Code], § 74, para. 2, https://www.gesetze-im-internet.de/englisch_hgb/ [https://perma.cc/RSK9-8DAU] (Ger.) (requiring employers to compensate employees during the term of the post-contractual non-compete at a minimum of fifty percent of most recent earnings).

141 *See id.*

142 *Navigating Non-Compete Clauses in Europe: A Comprehensive Guide*, GOGLOBAL (Sept. 16, 2024), https://goglobal.com/blog/employer-of-record/navigating-non-compete-clauses-in-europe-a-comprehensive-guide [https://perma.cc/LN2T-V64S].

judicial enforcement, foster voluntary compliance and reduce adversarial enforcement mechanisms.[143]

c. Comparative Perspective

In contrast, jurisdictions with categorical bans (e.g., California) promote worker mobility and innovation but may reduce incentives for employer investment in training due to limited post-employment protections.[144] Meanwhile, traditional common law systems that rely solely on a judicial reasonableness standard often produce inconsistent outcomes and higher litigation costs, due to their lack of statutory guidance or compensation requirements.[145] Germany's structured approach—combining enforceability with fairness mechanisms—yields a more balanced, predictable, and transparent system for managing post-employment competition risks.

2. Implementation Challenges

Despite the advantages of the German approach, adopting a similar model in the United States presents significant implementation challenges. Particularly, these challenges are complex due to the United States' decentralized legal system, variation in employment law across states, and the absence of a unified labor market enforcement agency. Two key categories of challenges emerge.

a. Administrative and Legal Infrastructure

The United States lacks a federal-level administrative body with the capacity or mandate to enforce non-compete restrictions uniformly. The implementation of a German-style framework would require the following:

- Federal Legislation or Coordinated State Action: Unlike Germany's centralized legal code, U.S. employment law is governed primarily at the state level. Instituting a nationwide framework would either require preemptive federal legislation or uniform model laws adopted across all states, which presents political and legal hurdles.[146]
- Specialized Adjudication: German labor courts have specific jurisdiction over employment disputes. In the

143 *See* Bodenstedt & Norda, *supra* note 137.

144 Rachel E. Green, *The Latest Attack on California's Ban of Noncompete Agreements*, KATZ BANKS KUMIN (June 21, 2024), https://katzbanks.com/employment-law-blog/california-noncompetes [https://perma.cc/EY8W-5S6G].

145 Starr, Prescott & Bishara, *supra* note 44.

146 *Id.*

United States, employment contract disputes are typically handled by general civil courts, which may lack subject-matter expertise in evaluating the reasonableness and enforceability of non-competes with compensation.[147]

- Monitoring and Enforcement Systems: There is no equivalent to Germany's local employment agencies (*Arbeitsagenturen*) or statutory compliance bodies. U.S. implementation would require creating or delegating regulatory oversight capacity—whether through the Department of Labor (DOL), FTC, or new state-level bodies.[148]
- Estimated Costs: According to U.S. Department of the Treasury (DOT) modeling, the administrative and legal infrastructure required to monitor enforceable non-competes with mandatory compensation could cost $1.3–$2 million per million workers covered. However, the DOT also found that such costs may be offset by reduced litigation and improved labor mobility within three to five years.[149]

b. Market Adaptation and Business Resistance

The transition to a German-style compensatory model would also present significant adaptation challenges for U.S. employers. Many U.S. employers—particularly in states like Florida, Texas, and New York—rely on non-competes as routine elements of employment contracts. A compensatory regime would represent a dramatic departure from both the common law "reasonableness" standard and the freedom-of-contract tradition.[150] Many U.S. employers would view the German model, under which enforceable non-competes must be compensated at fifty percent of the employee's most recent earnings, as an unfunded mandate, especially for roles involving large workforces or low-margin operations.[151]

Implementation would also require massive training programs, Human Resources (HR) system upgrades, and contract

147 *See generally* U.S. DEP'T OF THE TREASURY, NON-COMPETE CONTRACTS: ECONOMIC EFFECTS AND POLICY IMPLICATIONS 15, 28–30 (2016), https://home.treasury.gov/system/files/226/Non_Compete_Contracts_Econimic_Effects_and_Policy_Implications_MAR2016.pdf [https://perma.cc/7L2N-MCQ8].

148 *Non-Compete Clause Rulemaking*, FED. TRADE COMM'N (Jan. 5, 2023), https://www.ftc.gov/legal-library/browse/federal-register-notices/non-compete-clause-rulemaking [https://perma.cc/UC5J-25SP] (last visited Apr. 12, 2025).

149 U.S. DEP'T OF THE TREASURY, *supra* note 147, at 19–20, 20 tbl.1.

150 *See* Barnett & Sichelman, *supra* note 24, at 953–54.

151 *See Navigating Non-Compete Clauses in Europe*, *supra* note 142.

template redesigns across businesses, especially in mid-sized and small enterprises that lack in-house legal departments.[152] Sectors that rely heavily on human capital—like tech, finance, and pharma—would be more amenable to adopt and manage compliance. But lower-wage industries, where non-competes are already disproportionately imposed, may struggle to adapt without substantial legal reform and oversight.[153]

3. Transition Challenges and Success Factors in Adopting the German Model in the United States

Adopting a German-style compensatory system for non-compete agreements in the United States presents distinct and multifaceted challenges.

a. Specific Transition Challenges

Industries exhibit varying capacities to adapt to a compensatory non-compete model. Knowledge-intensive sectors (e.g., tech, finance, biotech) tend to adjust more readily due to pre-existing compensation structures, sophisticated HR and legal compliance systems, greater financial flexibility, and institutionalized training programs.

In contrast, traditional and low-margin sectors—including retail, manufacturing, and hospitality—may face significant transition barriers. For instance, there may be limited liquidity to fund mandatory compensation, underdeveloped internal compliance infrastructure, legacy employment practices, and managerial resistance to contractual change.[154]

Firm size strongly correlates with transition capacity. Large enterprises typically adapt more easily due to scale, internal legal teams, and robust systems already in place. Mid-sized firms often lack the same degree of HR or legal infrastructure and may require transitional support. Small businesses may face disproportionately high compliance burdens without assistance. Startups and high-growth firms may need tailored guidance to preserve agility while complying with compensation mandates. Multinational corporations face additional challenges due to jurisdictional

152 *See id.*

153 Alexander J.S. Colvin & Heidi Shierholz, *Noncompete Agreements*, ECON. POL'Y INST. (Dec. 10, 2019), https://www.epi.org/publication/noncompete-agreements/ [https://perma.cc/C8VJ-EPMP].

154 *See Navigating Non-Compete Clauses in Europe*, *supra* note 142.

coordination across countries with divergent legal norms.[155]

b. Critical Success Factors

Empirical research and international experience suggest that gradual rollout of non-compete reform increases the likelihood of sustainable adoption. This phased approach may begin with pilot programs in selected industries or regions, followed by expansion based on workforce segment or firm size.[156] Built-in evaluation intervals should be incorporated to assess progress at key stages.[157] Ongoing consultation with stakeholders is also essential to ensure responsiveness, along with regulatory flexibility to adjust based on results.[158]

In addition, comprehensive support systems are critical. Governments or coalitions of agencies may need to provide technical assistance to guide businesses on legal and operational compliance.[159] Financial support mechanisms, such as tax credits or phased-in obligations, may be necessary for small employers. The rollout should also be accompanied by standardized tools and resources, including model contract language and compliance toolkits.[160] Educational programs for HR managers, legal counsel, and employees and industry-specific implementation strategies play a key role in fostering understanding and adherence.[161]

Finally, measurable implementation goals are essential to track impact and build stakeholder trust. Common metrics include reduction in non-compete-related litigation, voluntary compliance rates by employer size and sector, and changes in employee wage trajectories.[162] Additionally, data on labor market mobility—such as the rate of job transitions within industries—should be monitored along with indicators of

---

155 *See Non-Compete Clauses in the UK and U.S.: Recent Trends*, *supra* note 132 (discussing the varied non-compete laws in the UK, the United States, and other European jurisdictions).

156 *See* Christopher Caiaccio et al., *A Comprehensive Update on Recent Federal and State Efforts to Limit the Use of Employee Non-Compete Agreements*, JD SUPRA, https://www.jdsupra.com/legalnews/a-comprehensive-update-on-recent-8977470 [https://perma.cc/9GE3-JEXV] (last visited Apr. 12, 2025).

157 *Id.*

158 *Id.*

159 *See Navigating Non-Compete Clauses in Europe*, *supra* note 142; *Non-Compete Clause Rulemaking*, *supra* note 148.

160 *See* sources cited *supra* note 159.

161 *See* sources cited *supra* note 159.

162 Evan Starr, *Noncompete Clauses: A Policymaker's Guide Through the Key Questions and Evidence*, ECON. INNOVATION GRP. (Oct. 31, 2023), https://eig.org/noncompetes-research-brief [https://perma.cc/S5K9-2YQP].

regional or sectoral innovation activity.[163]

I. Why the German Model-Based Proposal Is Superior

Unlike the Workforce Mobility Act, which eliminates non-compete agreements entirely,[164] the German model provides a structured compromise that preserves employer protections while ensuring worker fairness. The key advantages of a German-modeled proposal include the following:

1. Mandatory Compensation: Employers must compensate workers during the restricted period (typically at least fifty percent of salary), ensuring that workers are not left without income while being restricted from seeking employment elsewhere.
2. Tailored Restrictions: Instead of a blanket prohibition, non-competes are enforceable only if they meet clear, predefined criteria—such as reasonable duration (two years max), geographic scope, and industry-specific needs.
3. Judicial Oversight and Enforcement Mechanisms: Rather than banning agreements outright, a regulatory and judicial framework ensures that only justified agreements are upheld, discouraging abusive or overly broad restrictions.
4. Flexibility for Business and Workers: The German model allows businesses to retain key employees in highly sensitive roles while enabling workers to negotiate fairer exit terms. This system reduces litigation and administrative burdens, as companies must carefully consider whether imposing a non-compete is worth the financial cost.

By adopting a compensatory approach rather than a blanket ban, the German model avoids the pitfalls of both the Workforce Mobility Act's overcorrection[165] and the state-by-state inconsistency that currently exists in the United States.[166] The

163 *Id.*

164 Workforce Mobility Act of 2023, S. 220, 118th Cong. (2023).

165 *See generally* Kristopher Kalkowski, *Recognizing an Overcorrection: A Proposal for Nevada's Policy on Non-Compete Agreements*, 18 NEV. L.J. 261 (2017).

166 *State Noncompete Law Tracker*, ECON. INNOVATION GRP. (Oct. 11, 2024), https://eig.org/state-noncompete-map [https://perma.cc/S4V3-PE5N] (detailing which states have banned non-competes, which have either income-based or other restrictions, and which have no restrictions at all).

proposal allows for national uniformity while preserving economic incentives for innovation and workforce development.

Ultimately, a federal compensatory framework provides the best alternative to both extreme deregulation and excessive restriction. The Workforce Mobility Act, while well-intentioned, swings too far in one direction—just as the FTC's categorical ban does—whereas the German model-based approach balances business needs and worker rights in a practical, economically viable way.

V. ECONOMIC AND EMPIRICAL ANALYSIS OF NON-COMPETE AGREEMENTS

A. The Impact on Wages and Job Mobility

Numerous empirical studies indicate that non-compete agreements reduce wages and limit job mobility. A 2019 study by the Economic Policy Institute found that non-competes lower wages by five to ten percent in affected industries.[167] Employees bound by non-competes have been shown to change jobs less frequently, hindering wage growth and career advancement.[168]

In contrast, states that prohibit non-competes, such as California, experience higher wage growth, particularly in technology and knowledge-based industries.[169] The evidence suggests that restricting non-competes promotes a more dynamic labor market and encourages competition among employers.[170]

B. Effect on Innovation and Business Growth

Non-competes can stifle innovation by preventing skilled employees from launching startups or moving to competitors where they can contribute to technological advancements.[171] Silicon Valley's success is often attributed to the free flow of talent, in part due to California's ban on non-competes.[172]

However, businesses argue that non-competes protect proprietary knowledge and ensure return on investment in

---

167 Colvin & Shierholz, *supra* note 153 (table 4).

168 See Bhargav Gopal, Xiangru Li & Luke Rawling, Do Non-Compete Agreements Help or Hurt Workers? Evidence from the NLSY97, at 1 (Apr. 15, 2025) (unpublished research paper), https://bhargavgopal.com/resources/paper2.pdf [https://perma.cc/7UYQ-YHKS].

169 *See* Barnett & Sichelman, *supra* note 24, at 956–57, 1008.

170 *See generally* Erik Stam, *The Case Against Non-Compete Agreements* (Utrecht Univ. Sch. Econ., Working Paper No. 19-20, 2019), https://www.uu.nl/sites/default/files/rebo_use-wp_2019_1920.pdf [https://perma.cc/R49X-H24B].

171 *See* Barnett & Sichelman, *supra* note 24, at 953.

172 *See id.* at 956–57.

employee training.[173] A balanced system—such as the German model—which mandates compensation, ensures that businesses do not exploit non-competes while still protecting their interests.[174]

C. Industry-Specific Impacts

The effects of non-compete agreements vary significantly across industries. Some sectors rely heavily on these agreements to protect trade secrets and investments, while others see them as barriers to fair competition and worker mobility.[175]

1. Technology Sector

In the technology industry, non-compete agreements are commonly used to prevent employees from taking proprietary knowledge to competing firms. Employers argue that without these agreements, firms risk losing intellectual capital, leading to unfair competitive disadvantages.[176] However, empirical evidence suggests that non-compete bans, such as in California, have spurred innovation by allowing talent to move freely.[177] Silicon Valley's high concentration of tech firms, rapid knowledge sharing, and fast-paced innovation environment exemplify the benefits of reduced restrictions.[178]

2. Healthcare Industry

The healthcare sector has faced increasing scrutiny regarding non-compete clauses, particularly for physicians and specialists.[179] Many states have passed laws restricting or banning non-competes in healthcare due to concerns over patient access and continuity of care.[180] Studies have shown that physician non-competes can increase healthcare costs by limiting provider availability and creating regional monopolies where patients have fewer choices.[181] Additionally, restrictive agreements discourage

---

173 *See id.* at 970.

174 *See* discussion *infra* Section IV.I.

175 *See* Sampsa Samila & Olav Sorenson, *Noncompete Covenants: Incentives to Innovate or Impediments to Growth*, 57 MGMT. SCI. 425, 425–26 (2011).

176 *See* Barnett & Sichelman, *supra* note 24, at 970.

177 *See id.* at 956–57.

178 *See* Bruce Fallick, Charles A. Fleischman & James B. Rebitzer, *Job-Hopping in Silicon Valley: Some Evidence Concerning the Microfoundations of a High-Technology Cluster*, 88 REV. ECON. & STAT. 472, 472 (2006).

179 Kurt Lavetti, Carol Simon & William D. White, *The Impacts of Restricting Mobility of Skilled Service Workers: Evidence from Physicians*, 55 J. HUM. RES. 1025, 1026–28 (2020).

180 *See* J. Jeffrey Marshall et al., *Restrictive Covenants and Noncompete Clauses for Physicians*, 2 JACC: ADVANCES 1, 2 (2023).

181 *See* Kurt Lavetti, Carol Simon & William D. White, Buying Loyalty: Theory and Evidence from Physicians 1–5 (Oct. 26, 2012) (unpublished manuscript), https://www.sole-jole.org/assets/docs/13228.pdf [https://perma.cc/559H-BQ7H].

doctors from setting up independent practices, further consolidating power among large hospital systems.[182]

3. Manufacturing and Trade Industries

In the manufacturing industry, non-compete agreements often serve to protect specialized processes, customer relationships, and proprietary technology. While some firms rely on them to maintain competitive advantages, overly restrictive agreements can suppress worker mobility and limit wage growth.[183] Unlike technology firms that can use alternative protections (e.g., patent law), manufacturing firms frequently argue that non-competes are necessary due to the direct hands-on nature of their trade secrets.[184]

4. Financial and Professional Services

The financial and legal sectors frequently employ non-competes, particularly to prevent professionals from taking clients with them when changing firms. However, these agreements have led to debates about whether they truly protect firms or simply serve to suppress wages.[185] While protecting client lists may be a legitimate business concern, critics argue that non-competes often go beyond necessity, restricting professionals from fairly competing in the market.[186]

5. Retail and Low-Wage Employment

Non-compete clauses in retail and low-wage sectors have been widely criticized as exploitative.[187] Employers in industries like fast food, sales, and hospitality have used these agreements to prevent workers from moving to higher-paying jobs.[188] This practice has been labeled anti-competitive and detrimental to wage growth,

---

182 *See, e.g.*, Marshall et al., *supra* note 180, at 1.

183 *See* Colvin & Shierholz, *supra* note 153.

184 *See id.*

185 *See* discussion *supra* Sections V.A–V.B.

186 *See* Lavetti et al., *supra* note 179, at 5–6.

187 *See, e.g.*, Naomi Kodama, Ryo Kambayashi & Atsuko Izumi, *Non-Compete Agreements: Human Capital Investments or Compensated Wages?* 1 (IZA Inst. Lab. Econ., Discussion Paper No. 17685, 2025), https://docs.iza.org/dp17685.pdf [https://perma.cc/A8L9-2KZM]; Dave Jamieson, *Jimmy John's Makes Low-Wage Workers Sign 'Oppressive' Noncompete Agreements*, HUFFPOST (Oct. 13, 2014, 4:03 PM), https://www.huffpost.com/entry/jimmy-johns-non-compete_n_5978180?1413230622 [https://perma.cc/P5MG-MNCU].

188 *See* Jamieson, *supra* note 187.

leading states like Illinois[189] and Washington[190] to prohibit non-competes for workers earning below a certain salary threshold.

VI. PROPOSED FEDERAL FRAMEWORK FOR NON-COMPETE AGREEMENTS

A. Introduction: The Need for Federal Harmonization

As discussed, the current regulatory landscape surrounding non-compete agreements in the United States is fragmented, inconsistent, and increasingly unsustainable. While some states enforce non-competes liberally, others impose near-total bans. The result is a jurisdictional patchwork that creates uncertainty for employers, restricts employee mobility, and undermines national labor market fluidity. This variation in enforceability also encourages forum shopping, creates compliance burdens for multistate employers, and reinforces structural inequities in bargaining power. A federal framework is urgently needed—not to outlaw non-compete agreements outright, but to replace the current chaos with a coherent regulatory structure that promotes innovation, labor mobility, and legitimate business protection.

The proposal advanced in this section is inspired by the German model of regulated non-compete agreements, which is widely regarded as one of the most balanced systems internationally. Under German law, employers must provide post-employment compensation to enforce a non-compete clause, which fundamentally alters the employer's cost-benefit calculus.[191] Agreements must also meet strict proportionality, temporal, and geographic limitations. These principles inform the proposed U.S. framework, which seeks to regulate rather than prohibit non-compete agreements and to achieve equilibrium between worker freedom and business security.

B. Core Provisions of the Federal Framework

The proposed federal statute would set a uniform national standard for non-compete enforceability. States could enact stricter provisions but not weaker ones, much like the relationship

189 *See* Freedom to Work Act, 820 ILL. COMP. STAT. 90/10 (2025).

190 *See* WASH. REV. CODE § 49.62.020 (2025); James Sanders et al., *Washington State Tightens Noncompete Restrictions*, PERKINS COIE (May 2, 2024), https://perkinscoie.com/insights/update/washington-state-tightens-noncompete-restrictions [https://perma.cc/MQH3-RNWY].

191 Michael Magotsch & Pascal R. Kremp, *Non-Competition Clauses*, *in* KEY ASPECTS OF GERMAN EMPLOYMENT AND LABOUR LAW 161–70 (Jens Kirchner, Pascal R. Kremp & Michael Magotsch eds., 2d ed. 2018).

between federal and state minimum wage laws. The framework would include the following key elements.

1. Mandatory Compensation Requirement

A non-compete agreement would be enforceable only if the employer agrees to pay at least fifty percent of the employee's average total compensation (including bonuses, commissions, equity, and benefits) for the duration of the restriction period. This mirrors section 74a paragraph 2 of HGB, which requires compensation as a condition for enforceability.[192] The economic rationale is simple—if employers must pay to restrict post-employment conduct, they will use non-competes selectively and strategically. German empirical studies show that the introduction of mandatory compensation dramatically reduces the number of non-competes imposed without impairing business performance or innovation rates.[193]

2. Maximum Duration of Two Years

The restriction period may not exceed two years post-employment. This mirrors the cap found in HGB section 74a paragraph 1 and reflects the general consensus among U.S. courts that longer durations rarely satisfy reasonableness tests, especially for non-executive positions.[194] A two-year cap is also consistent with best practices in states like Massachusetts and Oregon, which impose similar maximum durations.[195] A bright-line rule would enhance legal clarity and reduce litigation costs.

3. Proportionality and Legitimate Business Interest Test

To be enforceable, a non-compete must be narrowly tailored to protect a legitimate business interest, such as trade secrets, proprietary processes, or key client relationships.[196] Vague justifications—like "preventing competition" in general—would be insufficient. Courts would apply a three-part proportionality test, assessing (1) the employer's business justification, (2) the scope

---

192 *Handelsgesetzbuch* [HGB] [Commercial Code], § 74, para. 2, https://www.gesetze-im-internet.de/englisch_hgb/ [https://perma.cc/RSK9-8DAU] (Ger.).

193 Magotsch & Kremp, *supra* note 191.

194 Norman D. Bishara, Kenneth J. Martin & Randall S. Thomas, *An Empirical Analysis of Noncompetition Clauses and Other Restrictive Postemployment Covenants*, 68 VAND. L. REV. 1, 39–41 (2015).

195 MASS. GEN. LAWS CH. 149 § 24L(b)(iv) (2024); OR. REV. STAT. § 653.295(3) (2024).

196 Kim A. Leffert, Andrew S. Rosenman & Ruth Zadikany, *United States: Restrictive Covenants*, MAYER BROWN (July 25, 2024), https://www.mayerbrown.com/en/insights/publications/2024/07/restrictive-covenants-us [https://perma.cc/4VDH-K8NE].

and duration of the restriction, and (3) the economic impact on the employee.[197] This approach is widely used in European labor law and would codify a structured reasonableness standard in the United States.[198]

4. Geographic and Activity Scope Limitations

The restriction must be limited to geographic areas where the employer actually conducts business, not merely areas of aspirational interest. Similarly, the activity restriction must be limited to the employee's actual role or exposure to sensitive information. This codifies existing case law that disfavors overly broad restraints and reflects a consistent line of German court rulings invalidating non-competes that attempt to cover global or undefined markets.[199]

5. Written Notice and Advance Disclosure

Employers would be required to provide written notice of the non-compete agreement at the time of the job offer, or at least fourteen days before the employment start date. Any agreement introduced after employment begins would require additional consideration, such as a raise or bonus. These procedural protections mirror disclosure requirements adopted in Illinois and Washington and are designed to ensure informed consent and procedural fairness.[200]

6. Exemptions and Special Treatment

The federal framework would exempt low-wage workers from non-compete enforcement entirely. Following models already adopted in states like Illinois and Washington, the statute would prohibit non-competes for employees earning below a defined threshold, indexed annually for inflation.[201] Special rules could also apply to executives, equity holders, and research and

197 *Id.*

198 *Id.*

199 Mahnhold, *supra* note 103, at 333–34.

200 820 ILL. COMP. STAT. 90/20 (2024); WASH. REV. CODE § 49.62.020(1)(a)(i) (2025).

201 *See* 820 ILL. COMP. STAT. 90/10(a)–(b) (setting income thresholds for non-compete agreements); WASH. REV. CODE § 49.62.020–.040 (establishing low-wage exemptions and providing for inflation adjustments).

development (R&D) personnel, allowing greater latitude in exchange for higher levels of compensation or severance.

C. Advantages of a Compensatory Approach

Unlike the categorical ban proposed by the FTC, this compensatory framework preserves freedom of contract, while discouraging overuse of non-compete agreements through market-based incentives. By requiring employers to pay to restrict worker mobility, the law encourages selective and strategic use, rather than reflexive overreach. This respects the autonomy of both parties and aligns with foundational principles of contract theory and labor economics.

Moreover, a compensatory approach aligns with the empirical evidence. Research from Germany, California, and other U.S. states with partial bans shows that when non-compete enforceability is restricted or made costly, job mobility increases, wages rise, and innovation flourishes—without significant increases in trade secret litigation.[202] These findings suggest that the competitive harms feared by employers are often overstated, and that firms adapt quickly to a more mobile workforce by strengthening internal retention tools and confidentiality practices.[203]

D. Administration, Enforcement, and Compliance Mechanism

To avoid regulatory ambiguity and litigation overload, the proposed statute would establish a centralized federal authority to implement and oversee the new rules governing non-compete agreements. The DOL would be the most appropriate agency to house this function through a new Office for Labor Market Fairness and Mobility, which would be granted rulemaking, investigative, and enforcement powers.[204]

1. Rulemaking Authority

The DOL would be authorized to issue interpretive guidance and administrative rules defining key statutory terms, such as "legitimate business interest," "confidential information," and "reasonable geographic scope." Drawing from administrative precedents in wage-and-hour law, the agency could

---

202 *See* Starr, Prescott & Bishara, *supra* note 44, at 53–84.

203 *See* Matt Marx, Jasjit Singh & Lee Fleming, *Regional Disadvantage? Employee Non-Compete Agreements and Brain Drain*, 44 RSCH. POL'Y 394, 394–404 (2015).

204 *See Paul DeCamp Featured in "Former DOL W&H Head Talks Shop on Agency Rulemaking,"* EPSTEIN BECKER GREEN (Aug. 27, 2024), https://www.ebglaw.com/insights/news/paul-decamp-featured-in-former-dol-w-h-head-talks-shop-on-agency-rulemaking [https://perma.cc/592G-3X7T].

create industry-specific standards that reflect variations in market structure, trade secret sensitivity, and typical employment durations. For example, the tech sector may receive more latitude on non-competition agreements, while the retail sector might face stricter scrutiny due to limited proprietary exposure.

2. Filing and Registration Requirement

To enforce a non-compete agreement, the employer must file the agreement with the DOL within ten business days of execution. The filing would include: (1) the employee's name and position; (2) the restricted activities and geographic scope; (3) the duration of the restriction; and (4) the compensation structure to be paid during enforcement. Failure to register the agreement would render it per se unenforceable. This registration requirement increases compliance while also creating a centralized, anonymized database for researchers, policymakers, and advocacy groups to monitor trends and identify potential abuses.

3. Complaint Mechanism and Whistleblower Protections

Employees would have access to a streamlined complaint process through the DOL's Wage and Hour Division, including online- and phone-based intake systems. To ensure fair access, the law would prohibit retaliation against employees who report violations or who refuse to sign unenforceable agreements. Whistleblower protections modeled on the Sarbanes-Oxley (SOX) and Dodd-Frank frameworks would prohibit retaliation against individuals who report violations, ensuring that employees can seek relief without risking their livelihoods.[205]

These frameworks—modeled on the Sarbanes-Oxley Act of 2002[206] and the Dodd-Frank Wall Street Reform and Consumer Protection Act of 2010[207]—offer robust, multi-layered protection for employee whistleblowers. SOX prohibits publicly traded companies from retaliating against employees who report conduct they reasonably believe constitutes fraud or a violation of federal securities law.[208] Protected activity includes internal reporting,

---

205 *See* Sarbanes-Oxley Act of 2002, Pub. L. No. 107-204, 116 Stat. 745 (codified as amended in scattered sections of 15 & 18 U.S.C.); Dodd-Frank Wall Street Reform and Consumer Protection Act, Pub. L. No. 111-203, 124 Stat. 1376 (2010) (codified as amended in scattered sections of 5, 7, 12, 15, 18 & 31 U.S.C.).

206 Sarbanes-Oxley Act of 2002, 18 U.S.C. § 1514A.

207 Dodd-Frank Wall Street Reform and Consumer Protection Act, 15 U.S.C. § 78u-6.

208 18 U.S.C. § 1514A(a)(1).

testifying, or assisting in investigations by regulatory agencies. [209] The statute provides for reinstatement, back pay, and special damages including attorney's fees and costs.[210] Dodd-Frank builds on SOX by providing a private right of action for whistleblowers who suffer retaliation, extending the statute of limitations to six years and offering remedies, such as double back pay with interest.[211] It also allows employees to report directly to the Securities and Exchange Commission (SEC), even bypassing internal reporting channels.[212]

SOX and Dodd-Frank, particularly when combined, represent robust federal whistleblower protection. Incorporating their core features into the proposed non-compete enforcement framework would ensure that workers, especially those in vulnerable or low-wage positions, can report violations and coercive practices without fear of job loss, blacklisting, or other forms of employer retaliation.

E. Oversight, Sunset Clause, and Empirical Evaluation

One of the flaws in previous regulatory attempts—including the FTC's 2024 non-compete rule—is the lack of an embedded feedback loop. To remain responsive to evolving labor markets, the federal framework would include robust oversight and revision mechanisms, including a statutory sunset clause.

1. Biennial Impact Review

The DOL, in partnership with the Bureau of Labor Statistics and external academic institutions, would conduct biennial reviews. The reviews will include assessments of (1) the number and types of non-competes filed; (2) litigation rates; (3) impacts on wage growth, job switching, and innovation; and (4) sector-specific enforcement patterns. Once conducted, these reviews would be publicly released and subject to notice-and-comment procedures to allow stakeholder feedback. The review process mirrors successful federal initiatives in occupational

209 18 U.S.C. § 1514A(a)(1)–(2); *see also* Sylvester v. Parexel Int'l LLC, ARB Case No. 07-123, 2011 WL 2165854, at *27 (Dep't of Lab. May 25, 2011) (broadly construing scope of protected activity under SOX).

210 18 U.S.C. § 1514A(c)(2).

211 15 U.S.C. § 78u-6(h)(1)(B).

212 *See* Digit. Realty Tr., Inc. v. Somers, 583 U.S. 149, 152 (2018) (holding that Dodd-Frank whistleblower protections apply only to individuals who report directly to the SEC).

safety and public health that are driven by continuous data gathering and recalibration.

2. Ten-Year Sunset Clause with Reauthorization

To prevent regulatory ossification, the statute would include a ten-year sunset provision. If not reauthorized by Congress after a formal review, the law would expire—unless the DOL certifies that its continuation is necessary and effective. This sunset clause would create political accountability while building in structural flexibility, a design choice increasingly common in major economic legislation.

F. Addressing Industry-Specific Concerns

The most persistent criticism of non-compete regulation is its potential to undermine trade secret protection and weaken industry-specific competitive advantages. While much of this criticism is overstated—especially in low-skill sectors—it is not without merit in certain high-risk or high-investment fields.

1. R&D-Heavy and IP-Driven Industries

Sectors like biotechnology, defense, pharmaceuticals, and advanced manufacturing often involve long product cycles, large R&D expenditures, and deep knowledge capital embedded in key employees. In these contexts, the employer's competitive edge may hinge on preventing defectors from joining rival firms with highly substitutable product pipelines.[213]

To address these concerns, the proposed framework allows for the following heightened protection measures:

- The compensation floor (fifty percent) could be increased voluntarily by employers in exchange for longer duration or broader scope, subject to DOL oversight.
- Executive roles and key intellectual property positions could be subject to enhanced non-disclosure and non-solicitation provisions, enforceable independently of the non-compete clause.
- In rare cases, employers could petition the DOL for a waiver of the two-year cap, with rigorous justification

213 ORLY LOBEL, TALENT WANTS TO BE FREE 134–39 (2013).

based on product development timelines or national security sensitivity.

These safeguards strike a balance—they allow flexibility for firms whose survival depends on proprietary information, but require transparency, compensation, and agency oversight.

2. Small Businesses and Startups

Critics of regulation often point out that small firms and startups may rely disproportionately on non-competes due to their limited resources for formal intellectual property enforcement. However, a compensatory model addresses this concern by leveling the playing field—firms will use non-competes only when truly necessary and when they can afford to do so. Moreover, confidentiality agreements and trade secret law already provide strong protection without burdening labor mobility.

Startups may also benefit from increased labor fluidity, as workers gain more flexibility to move between early-stage ventures. Empirical studies have shown that high-growth entrepreneurship is positively correlated with non-compete unenforceability.[214] Thus, the proposed framework may empower rather than constrain innovation ecosystems.

3. Low-Wage and Franchise Employment

In the current environment, non-competes are still widely used in retail, food service, and personal care sectors, often with no legitimate justification. These agreements are used more to deter turnover than to protect legitimate competitive interests. The proposed framework would prohibit enforcement of non-competes for any employee earning less than the annually adjusted threshold, eliminating the most egregious and exploitative uses of these contracts.

Furthermore, franchisors would be prohibited from imposing blanket non-compete requirements on franchisee employees unless they can demonstrate access to proprietary data, training, or recipes. This responds to widespread criticism of cases like

214 *See, e.g.*, Sampsa Samila & Olav Sorenson, *Noncompete Covenants: Incentives to Innovate or Impediments to Growth*, 57 MGMT. SCI. 425, 425 (2011).

Jimmy John's and other fast-food chains that used non-competes to restrict sandwich-makers and delivery drivers.[215]

G. Normative Foundations and Constitutional Harmony

A compensatory, regulatory model for non-compete agreements is not only efficient and empirically grounded—it also aligns with core American constitutional and legal values. Critics of federal regulation often invoke the sanctity of freedom of contract, yet that very principle supports a model in which both parties to a restrictive covenant understand and internalize the full economic consequences of their agreement.

Unlike categorical bans, which foreclose private ordering, a compensatory framework respects party autonomy while simultaneously correcting for power imbalances and information asymmetries in employment negotiations. The proposal does not prohibit non-competes; it simply conditions their enforceability on fairness and transparency, ensuring that employers bear the financial cost of post-employment restrictions rather than externalizing them to workers.

Furthermore, this approach poses minimal constitutional risk. It does not compel speech or conduct, nor does it eliminate contractual freedom outright. By operating through the Commerce Clause—regulating instruments that materially affect interstate labor markets—it falls squarely within established federal authority. In contrast, the FTC's ban was challenged for exceeding the scope of rulemaking powers under the FTC Act, a vulnerability this framework avoids by grounding its authority in direct congressional legislation.

H. Complementary Role of State Law

Notably, this proposed statute is not intended to displace all state-level regulation. Rather, it would function as a federal floor, setting minimum standards for enforceability while allowing states to adopt stricter protections if they choose. In this way, it resembles the FLSA[216] or Title VII of the Civil Rights

215 *See* Richard L. Hathaway, *Fast Food Non-Compete Agreements?*, KANE RUSSELL COLEMAN LOGAN (Jan. 7, 2015), https://www.krcl.com/insights/fast-food-non-compete-agreements [https://perma.cc/4946-LFB8].

216 *See supra* note 66.

Act,[217] which establish baseline rights but permit local innovation and augmentation.

For example, states like California, which already prohibit most non-competes, could maintain that approach. However, in jurisdictions where non-competes are still widely enforced—including many Southern and Midwestern states—the federal floor would provide critical protection for workers who currently lack bargaining power or legal recourse. Uniform baseline rules would also benefit multi-state employers, who currently face a compliance burden navigating multiple inconsistent standards.

This dual-sovereignty model not only enhances practical enforceability—it also insulates the statute from Dormant Commerce Clause challenges, since it does not mandate uniformity but merely establishes minimum rights.

I. Conclusion: From Doctrine to Practice

This proposed federal framework is a practical middle path between deregulated inconsistency and blanket prohibition. It ensures that non-compete agreements serve their legitimate purpose—protecting trade secrets and competitive investments—without being weaponized to suppress wages, deter job switching, or entrench employer power.

By mandating compensation, limiting duration and scope, and requiring advance disclosure, the framework creates a labor market equilibrium in which restrictions are rare, rational, and remunerated. It aligns legal doctrine with economic reality, drawing from successful international models while remaining faithful to American legal principles. Most crucially, it

217 Title VII prohibits employment discrimination based on race, color, religion, sex, and national origin. Title VII of the Civil Rights Act of 1964, 42 U.S.C. § 2000e-2(a). *See generally* David A. Garcia*, Title VII Does Not Preempt State Regulation of Private Club Employment Practices*, 34 HASTINGS L.J. 1107 (1983) (affirming the notion that Title VII sets a floor, not a ceiling, on anti-discrimination protections). The statute contains a provision stating that nothing in the title shall be construed to invalidate or limit any law of any state or political subdivision that provides equal or greater protection to individuals. 42 U.S.C. § 2000e-7. This clause ensures that Title VII serves as a minimum standard, allowing states to offer broader anti-discrimination protections.

creates a structure that is both administratively feasible and politically defensible.

VII. IMPLEMENTATION AND POLICY CONSIDERATIONS

A. Introduction: From Concept to Enforcement

Even the most thoughtfully constructed federal framework for non-compete agreements will fail if it is not effectively implemented. As the previous section proposed, a compensatory structure aligned with German labor law balances flexibility and fairness.[218] Yet enacting such a regime in the United States requires careful consideration of institutional capacities, constitutional authority, transitional logistics, and stakeholder dynamics.

This section explores the practical challenges and strategic opportunities involved in operationalizing a federal non-compete policy. It addresses questions of administrative enforcement, legal defensibility, judicial review, state preemption, and political feasibility. It aims to move the conversation from what the law should be to how it could and should be implemented at scale.

B. Choosing the Right Administrative Home

1. The Department of Labor as Primary Regulator

The most logical agency to administer a national non-compete regime is the DOL, given its existing authority over wage-and-hour enforcement, workplace fairness, and labor protections. The Wage and Hour Division already maintains infrastructure for field investigations, regulatory rulemaking, and complaint processing—all of which are necessary to enforce non-compete laws effectively. Unlike the FTC, whose recent attempt to ban non-competes was struck down for exceeding its statutory mandate,[219] the DOL operates squarely within the labor and employment sphere authorized by Congress.

Locating non-compete oversight within the DOL would allow rulemaking to be harmonized with other federal employment policies, such as the FLSA and the National Labor Relations Act.[220] It would also enable industry-specific guidance, updated wage thresholds, and tailored exemptions based on evolving labor

218 *See supra* Section VI.I.

219 *See supra* notes 4–9 and accompanying text.

220 *See* Fair Labor Standards Act, 29 U.S.C. §§ 201–219; National Labor Relations Act, 29 U.S.C. §§ 151–169.

market data. Crucially, the DOL's integration into the existing federal labor enforcement ecosystem would streamline compliance and allow cross-referencing with wage, hour, and classification violations already within the agency's purview.

2. A New Division for Contract Oversight

To ensure focus and specialization, Congress should authorize the creation of a Contract Fairness and Mobility Division within the DOL. This division would be responsible for maintaining a centralized registry of enforceable non-compete agreements, reviewing agreements for statutory compliance, issuing interpretive guidance and technical assistance to employers and courts, investigating violations, and referring serious or willful cases for civil enforcement actions.

Moreover, the division could adopt a tiered structure with a review office to administer agreement filings and issue advisory opinions; a compliance office to handle employer outreach and policy interpretation; and an enforcement unit empowered to audit firms, impose civil penalties, and coordinate with state attorneys general. The model here could resemble that of the Occupational Safety and Health Administration (OSHA), which operates through a blend of rulemaking, technical assistance, and strategic inspection programs.[221]

Additionally, the division should offer a preclearance process for employers to voluntarily submit draft agreements in advance of employee signature. This mirrors the SEC's comment letter process for securities filings and would encourage employers to shape their contracts in accordance with clear statutory standards. Preclearance could be incentivized by granting safe harbor status to approved agreements or reduced penalties for firms that submit them in good faith.

C. Judicial Review and Constitutional Considerations

1. Authority Under the Commerce Clause

A federal statute regulating non-compete agreements would likely pass constitutional muster under Congress's power to regulate interstate commerce. In today's labor market where remote work, digital infrastructure, and national firms dominate, employee mobility is not confined by state borders. Non-compete

221 U.S. DEP'T OF LAB., OSHA FACT SHEET: OSHA COMPLIANCE ASSISTANCE 1–2 (2009), https://www.osha.gov/sites/default/files/publications/compliance-assistance-factsheet.pdf [https://perma.cc/W4H3-GVAN].

agreements directly affect the movement of labor and the flow of knowledge across states, especially in knowledge-based industries.

Judicial precedent is strongly supportive of congressional authority over labor-related conduct with substantial interstate effects. For instance, Congress has successfully legislated in areas involving minimum wages, workplace safety, and employee classification, all under the Commerce Clause.[222] Given this history, there is little doubt that a law governing the enforceability of post-employment restraints—which restrict mobility across state lines—would satisfy the substantial effects test. Under the Commerce Clause, Congress has the authority to regulate activities that substantially affect interstate commerce. This principle was affirmed in *United States v. Darby Lumber Co.*,[223] where the Supreme Court upheld the FLSA, recognizing that labor conditions have a significant impact on interstate commerce. Further, in *Garcia v. San Antonio Metropolitan Transit Authority*, the Court confirmed that Congress could apply the FLSA to state and local governments, emphasizing that the regulation of employment terms falls within the scope of the Commerce Clause when such employment substantially affects interstate commerce.[224] Post-employment restraints, like non-compete agreements, can restrict an individual's ability to work across state lines, thereby affecting the national labor market and interstate commerce. Given that such restraints can hinder labor mobility and economic competition, a federal law regulating their enforceability would likely be considered a valid exercise of Congress's Commerce Clause powers under the substantial effects test.

Moreover, the national scope of the problem provides a rational basis for federal regulation. Studies have demonstrated how restrictive covenants depress innovation and suppress wages at a macroeconomic level.[225] This economic drag constitutes a classic justification for congressional intervention.

---

222 *See* United States v. Darby, 312 U.S. 100, 109, 115 (1941) (upholding the FLSA as a valid exercise of Congress's Commerce Clause power and affirming that regulation of labor standards, including wages and hours, is constitutional where such conduct substantially affects interstate commerce); Indus. Union Dep't v. Am. Petroleum Inst., 448 U.S. 607, 613–15 (1980) (recognizing that Congress may regulate workplace safety under the Commerce Clause by delegating authority to agencies like OSHA to issue rules affecting labor conditions in industries engaged in interstate commerce); Tony & Susan Alamo Found. v. Sec'y of Lab., 471 U.S. 290, 296–97, 306 (1985) (holding that the FLSA applies to employees engaged in activities that affect interstate commerce, even when employed by a nonprofit organization).

223 *Darby*, 312 U.S. at 109–10.

224 Garcia v. San Antonio Metro. Transit Auth., 469 U.S. 528, 530 (1985).

225 *See* KRUEGER & POSNER, *supra* note 42, at 4–6.

2. Avoiding the Pitfalls of Agency Overreach

Unlike the FTC's rule, which attempts to ban non-competes through agency rulemaking and was ultimately enjoined by a federal court, a statute passed by Congress and administered by the DOL would rest on far more solid legal ground. Critics of the FTC approach argued that the agency lacked clear statutory authority to adopt rules of such economic significance. These separation-of-powers concerns derailed the rule before it could be implemented.

By contrast, a DOL-administered statute would be built on express congressional authorization. Rather than categorically banning agreements, it would create conditions for enforceability—such as compensation and duration limits. This "earned enforceability" approach is more consistent with traditional labor policy and would withstand judicial scrutiny better than a sweeping ban.

Importantly, because the DOL already administers laws like the FLSA and the Family and Medical Leave Act (FMLA), it is an institution accustomed to interpreting broad statutory standards and updating regulations over time. Its rulemaking would benefit from established experience and could withstand post-*Chevron* scrutiny,[226] so long as it stays within congressional parameters.

3. Dormant Commerce Clause and Federalism

A well-drafted, federal non-compete law would avoid Dormant Commerce Clause[227] problems by creating uniformity and reducing burdens on interstate commerce. The Dormant Commerce Clause prohibits states from enacting legislation that discriminates against or unduly burdens interstate commerce.[228] The current patchwork of state noncompete laws forces multistate employers to navigate conflicting standards, increasing compliance costs and discouraging cross-border employment. This fragmented approach

226 *See* Chevron U.S.A., Inc. v. Nat. Res. Def. Council, Inc., 467 U.S. 837, 843–44 (1984) (holding that courts must defer to a federal agency's reasonable interpretation of an ambiguous statute), *overruled by* Loper Bright Enters. v. Raimondo, 603 U.S. 369 (2024).

227 The Dormant Commerce Clause is a judicially created doctrine inferred from the Commerce Clause, prohibiting states from enacting legislation that discriminates against or unduly burdens interstate commerce, even in the absence of federal regulation. *See* U.S. CONST. art. I, § 8, cl. 3. It reflects the principle that interstate economic activity must remain free from protectionist or inconsistent state laws that disrupt the national market. *See* Granholm v. Heald, 544 U.S. 460, 472 (2005) ("State laws that discriminate against interstate commerce face 'a virtually *per se* rule of invalidity.'") (quoting Philadelphia v. New York, 437 U.S. 617, 624 (1978)).

228 *Granholm*, 544 U.S. at 472.

creates significant friction in the national labor market and may amount to an undue burden on interstate commerce.

Because the Dormant Commerce Clause reserves regulatory authority over interstate commerce to Congress, Congress may invoke its power to reduce the burdens imposed by inconsistent noncompete laws. The Supreme Court has affirmed that Congress may regulate even local activities if they have a substantial effect on interstate commerce in the aggregate.[229] While the Dormant Commerce Clause generally protects against direct state-imposed burdens on interstate commerce, it does not invalidate all incidental burdens. [230] However, even facially neutral state laws must be invalidated when the burdens they impose on interstate commerce are clearly excessive in relation to their local benefits.[231]

States currently vary widely in how they treat non-competes, leading to uncertainty for employers operating across borders. A uniform baseline would reduce this friction, promote consistent labor standards, and improve compliance.

Federal preemption could be structured as a floor rather than a ceiling, allowing states like California and Washington to maintain stricter rules while ensuring that no state allows abusive enforcement below the federal floor.[232] This cooperative federalism model mirrors other federal labor statutes, including the Employee Retirement Income Security Act and the FMLA, which preempt some aspects of state law while preserving room for more protective measures.

D. Transitional Design and Implementation Timeline

1. Phase-In Period

Policymakers must design a measured transition that allows employers and workers to adapt to the new legal regime. A minimum eighteen-month phase-in period would provide time for employers to review and revise current agreements, for courts to receive interpretive guidance, and for state regulators to reconcile overlapping statutes. During this period, the DOL could issue interim rules, publish guidance documents, and offer

---

229 United States v. Lopez, 514 U.S. 549, 561 (1995).

230 Note, *The Dormant Commerce Clause and Moral Complicity in a National Marketplace*, 137 HARV. L. REV. 980, 983 (2024).

231 R. Randall Kelso, *The Proper Structure of Dormant Commerce Clause Review*, 59 TULSA L. REV. 109, 122 (2024).

232 *See* Chamber of Com. of the U.S. v. Bonta, 62 F.4th 473, 478 (9th Cir. 2023) (finding that California's Assembly Bill 51, a statute that attempted to bar the use of arbitration agreements by employers, is preempted by the Federal Arbitration Act).

employers a safe harbor for voluntary disclosures or good faith compliance efforts.

This phased approach would prevent mass invalidation of agreements and avoid placing businesses in sudden non-compliance. Transition rules should also recognize existing agreements entered into under good faith reliance on state law and offer "grandfathering" treatment where appropriate, subject to compliance with new compensation requirements. The Affordable Care Act offers a strong precedent for implementing labor reforms through phased rollout with delayed penalties and grace periods.[233]

2. Public Education and Compliance Outreach

Education and awareness will be critical to success. Workers often sign non-competes without legal counsel or full understanding of their implications. The DOL must fund a broad outreach initiative, including targeted online content and printed guides, worker rights materials in multiple languages, and partnerships with state agencies, bar associations, and labor unions.

Employers will also need technical assistance to revise contract templates, assess who qualifies for enforceable restrictions, and calculate proper compensation. The DOL should maintain a helpline, host webinars, and issue model clauses. Public access to the registry of enforceable agreements will deter illegal use and enable peer benchmarking.

E. Political Feasibility and Legislative Strategy

Implementing this framework in the United States would require navigating state autonomy and federal authority, but it offers a structured and equitable model guiding meaningful control. Although this model may face resistance due to increased costs and state-level control, the German framework encourages responsible use while preserving worker mobility.

1. Building a Coalition of Support

The political path to passage will require an unusual but achievable coalition. Worker advocates—particularly in labor unions and progressive think tanks—have long argued that

---

233 Affordable Care Act, 42 U.S.C. § 18001 (establishing the statute's implementation timelines).

non-competes suppress wages and hinder job mobility. These groups will champion any serious effort to curb abuse.

What makes this proposal distinctive is its potential appeal to employers as well. By avoiding a blanket ban, it preserves flexibility for firms that truly need post-employment protection. Startups, regional hospitals, manufacturers, and financial institutions would all prefer clear, enforceable standards to total prohibition. The law would give them a lawful path to retaining talent—at a cost, yes, but a known and manageable one.

Furthermore, a compensatory structure would appeal to free market advocates who favor internalizing costs over government bans. It also resonates with individuals interested in reducing litigation and clarifying legal obligations. With polling consistently showing that most workers disapprove of non-competes, and with evidence showing they harm wage growth and entrepreneurship,[234] the politics are increasingly favorable.

2. Anticipating Opposition and Messaging

Despite this broad potential support, opposition is to be anticipated from trade associations and some corporate lobbies that favor the status quo. These associations and lobbies will argue that any federal action preempts states, increases compliance burdens, and invites litigation. To counter this perspective, lawmakers must emphasize that the statute does not ban non-competes, but *conditions* them, preserving contractual freedom while elevating fairness.

Messaging should focus on *earned enforceability*: employers can still use non-competes, but only if they pay for the restriction and ensure the agreement is narrow in scope. It is a market solution, not a mandate. Moreover, states like Massachusetts and Oregon have already demonstrated that requiring notice and compensation reduces litigation and clarifies enforceability.[235]

F. Economic Risk Mitigation and Regulatory Design

1. Reducing Litigation Through Clarity

One of the strongest advantages of a compensatory federal framework is its capacity to reduce costly, unpredictable litigation.

---

234 Matthew S. Johnson, Kurt J. Lavetti & Michael Lipsitz, *The Labor Market Effects of Legal Restrictions on Worker Mobility* 36, 40 (Nat'l Bureau Econ. Rsch., Working Paper No. 31929, 2020), https://www.nber.org/system/files/working_papers/w31929/w31929.pdf [https://perma.cc/Y78P-QYKZ].

235 *See, e.g.*, MASS. GEN. LAWS ch. 149, § 24L (2024).

Current disputes over non-compete enforceability hinge on fact-intensive judicial tests of reasonableness in scope, geography, and duration. This common law patchwork has produced inconsistent rulings and incentivized employers to overreach, knowing litigation can chill worker mobility regardless of merit.[236]

By contrast, a federal standard could include clear statutory thresholds: for instance, a two-year maximum duration, income floors, and mandatory post-employment compensation. These bright-line rules would reduce reliance on vague multi-factor tests and help judges decide enforceability without expensive discovery or conflicting expert testimony.

Moreover, a registry of enforceable non-competes, filed with the DOL and available for public audit, would add transparency and deter abuse. Firms would be less likely to include unenforceable provisions if they know they must disclose terms in a federal filing accessible to regulators, courts, and competitors.[237]

2. Supporting Innovation Through Predictability

Critics of non-compete regulation often argue that restrictions are necessary to protect proprietary information and incentivize investment in employee training. But these concerns can be addressed through other mechanisms. Confidentiality agreements, intellectual property protections, and equity-based compensation remain enforceable and widely used.

Importantly, jurisdictions that restrict or prohibit non-competes—such as California—have not seen a collapse in innovation. In fact, venture capital formation, startup creation, and inventor mobility are consistently higher in these areas compared to enforcement-heavy states.[238]

The federal proposal does not prohibit non-competes; it requires employers to internalize the cost of restricting labor mobility. This market-based incentive structure forces firms to use non-competes selectively, reducing overuse and focusing attention on high-value, defensible cases. Structured compensation transforms non-competes from a risk-free default into a deliberate choice—supporting innovation by preserving talent circulation

236 *See* Barnett & Sichelman, *supra* note 24, at 988–90.

237 Nickolaus Stumo-Langer, Formerly Employed Need Not Apply 5 (Apr. 30, 2020) (unpublished manuscript), https://conservancy.umn.edu/server/api/core/bitstreams/fb3726a3-a86f-4e13-9211-a75ab213cc78/content [https://perma.cc/T679-QE6H].

238 Johnson, Lavetti & Lipsitz, *supra* note 234, at 16.

and knowledge spillovers.[239]

G. Leveraging Empirical Monitoring and Policy Feedback

1. Built-in Oversight and Adjustment

Effective implementation must be paired with ongoing evaluation. Congress should mandate that the DOL produce biennial reports tracking: (1) aggregate usage of non-competes by sector and geography; (2) average compensation offered for enforceable agreements; (3) complaint rates and litigation outcomes; and (4) worker mobility and wage growth metrics. This data will inform agency rulemaking and guide future legislative amendments. To ensure independent oversight, a National Non-Compete Observatory could be established in partnership with academic institutions and research foundations.

Additionally, the law should include a ten-year sunset clause, requiring Congress to review outcomes and reauthorize the statute based on empirical performance. This sunset model encourages legislative engagement and ensures the law evolves with labor market conditions.

2. Studying Sectoral and Regional Variability

The economic effects of non-competes vary widely across industries. High-tech fields, for example, are more sensitive to talent mobility, while manufacturing sectors often rely on specific process know-how and client relationships. Similarly, rural and urban labor markets may respond differently to restrictions based on density and competition.

The statute should therefore require data collection that is disaggregated industry (e.g., software, healthcare, finance); region (e.g., rural vs. metropolitan); and employer size (e.g., under 50 employees vs. Fortune 500).

This level of granularity will support more nuanced enforcement and future exemptions or modifications based on real-world effects. Without it, policymakers risk overcorrecting in one direction or failing to address new forms of abuse as they emerge.

H. Balancing Protection and Mobility in Practice

The long-standing tension between protecting legitimate business interests and preserving worker freedom is not new. But it

---

239 *See* Rohit Chopra & Lina M. Khan, *The Case for "Unfair Methods of Competition" Rulemaking*, 87 UN. CHI. L. REV. 357, 358, 373–74 (2020).

is newly urgent in a labor market shaped by technological disruption, widespread job switching, and growing economic inequality.

The proposed federal framework does not eliminate non-competes. It allows them—but only when employers can show that restrictions are narrow, justified, and paired with meaningful compensation. This earned enforceability model will discourage overuse, reduce litigation, and promote innovation by facilitating responsible use of non-competes.

Implementation matters as much as design. By situating enforcement in the DOL, building data systems, and adopting a phased rollout, the federal government can ensure a fair, functional, and future-proof regulatory structure. The goal is not to tip the scales—but to rebalance them.

VIII. ADDRESSING POTENTIAL CRITICISMS AND IMPLEMENTATION CHALLENGES

The proposed federal framework for non-compete agreements will likely face several significant criticisms and implementation challenges. This section addresses these potential objections while offering solutions to anticipated implementation difficulties.

A. Economic Impact Concerns

1. Business Cost Arguments

Critics will likely argue that mandatory compensation requirements impose excessive costs on businesses, particularly small and medium-sized enterprises.[240] This criticism warrants careful consideration but ultimately proves unpersuasive. Empirical evidence from Germany demonstrates that compensation costs are largely offset by reduced litigation expenses,[241] better employee retention, improved knowledge protection, enhanced innovation outcomes, and more efficient resource allocation. Additionally, the framework includes several cost-mitigation mechanisms, such as phased implementation schedules, small business assistance programs,

---

240 Evan Starr, Comment by Evan Starr to the Federal Trade Commission Re: Non-Compete Clause Rulemaking, Matter No. P201200, at 11 (Apr. 19, 2023) (unpublished comment), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4427741 [https://perma.cc/6QFC-LQSR].

241 KURT LAVETTI, NONCOMPETE AGREEMENTS IN EMPLOYMENT CONTRACTS 3, 5, 8–10 (2021), https://wol.iza.org/uploads/articles/578/pdfs/noncompete-agreements-in-employment-contracts.pdf [https://perma.cc/JS2L-YLR6].

tax incentives for compliance, industry-specific adjustments, and hardship exemptions.

2. Market Efficiency Arguments

Some critics may contend that federal regulation will reduce market efficiency by limiting employers' flexibility.[242] However, evidence suggests that structured regulation actually enhances market efficiency. This efficiency results from better information flow in labor markets due to transparent compensation requirements, clear restriction parameters, standardized agreement terms, predictable enforcement, and enhanced mobility data.

3. Innovation and Competition Arguments

Critics might argue that increased regulation of non-competes will stifle innovation and reduce competitive advantage.[243] In reality, empirical evidence suggests the opposite effect. First, structured compensation requirements actually promote innovation by encouraging targeted protection of valuable information, facilitating efficient knowledge transfer, supporting strategic R&D investments, promoting collaborative innovation, and reducing defensive patents.[244] Second, competitive advantages are enhanced through better alignment of restrictions with business needs, more efficient allocation of protection resources, reduced employee poaching costs, enhanced knowledge management, and improved succession planning.

4. Regulatory Burden Arguments

Another significant criticism concerns the administrative burden of compliance.[245] These concerns can be addressed through several mechanisms. The framework includes burden-reduction features, such as standardized documentation requirements, automated compliance systems, electronic filing platforms, streamlined reporting processes, and integration with existing HR

---

242 *See* Alan J. Meese, *Don't Abolish Employee Noncompete Agreements*, 57 WAKE FOREST L. REV. 631, 650–51 (2022).

243 Barnett & Sichelman, *supra* note 24, at 962–64.

244 *See generally* Karen Elliott et al., *Knowledge Protection in Firms: A Conceptual Framework and Evidence from HP Labs*, 16 EUR. MGMT. REV. 179 (2019); *see also* Jonathan M. Barnett, *Private Protection of Patentable Goods* 3 (Fordham Univ. Sch. L., Research Paper No. 28, 2003), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=445380 [https://perma.cc/KGA6-ZR8W].

245 Cameron Misner, Note, *Dependent Contractors? The Case for Giving Non-Competes a Central Role in Worker-Classification Tests Under Federal Law*, 109 CORNELL L. REV. 763, 770–71 (2024).

systems. Implementation assistance programs would provide technical support resources, compliance templates, training materials, industry-specific guidance, and small business assistance.

B. Federalism and Institutional Concerns

1. Federalism Challenges

The proposed framework must address significant federalism concerns regarding federal intervention in traditional state employment law.[246] State sovereignty considerations require careful attention, including traditional state regulation of employment, local economic conditions, state court expertise, existing state frameworks, and interstate competition. The framework addresses these concerns through cooperative federalism mechanisms, state implementation flexibility, local condition accommodation, existing institution utilization, and state-federal coordination.

2. Institutional Capacity Arguments

Critics may question whether federal institutions possess adequate capacity to implement comprehensive non-compete regulation. This challenge requires addressing several institutional components.

a. Administrative Capacity

The framework builds administrative capacity through specialized agency divisions, trained personnel development, technical infrastructure investment, interagency coordination, and state agency partnerships. Implementation would proceed through phased capacity building, pilot program testing, incremental jurisdiction expansion, regular capacity assessment, and resource allocation adjustment.

b. Judicial Capacity

The effectiveness of the proposed framework depends significantly on judicial capacity to handle non-compete disputes. Specialized judicial training programs would focus on non-compete agreement analysis, compensation calculation methods, geographic restriction evaluation, industry-specific considerations, and economic impact assessment. Structural modifications would enhance efficiency through specialized court

246 *See* Ernest A. Young, *Federal Preemption and State Autonomy*, *in* FEDERAL PREEMPTION: STATES' POWERS, NATIONAL INTERESTS 249, 249–76 (Richard A. Epstein & Michael S. Greve eds., 2007).

divisions, streamlined procedures, alternative dispute resolution, expert magistrates, and technical support staff.

3. Enforcement Coordination

Effective implementation requires sophisticated coordination among multiple enforcement entities. The framework establishes clear coordination mechanisms.

a. Vertical Coordination

The system provides for coordinated enforcement in a hierarchical structure between federal oversight agencies, state labor departments, local enforcement units, administrative tribunals, and traditional courts. Coordination would occur through joint enforcement protocols, information sharing systems, jurisdictional guidelines, resource sharing agreements, and unified compliance databases.

b. Horizontal Coordination

The framework also facilitates coordination among entities operating at the same enforcement level, including multiple federal agencies, different state authorities, various court systems, industry regulators, and professional associations.

C. Implementation Challenges and Practical Solutions

1. Transition Management

The shift to a federal compensatory system presents significant transition challenges that require careful management. The framework addresses these challenges through structured implementation phases.

a. Initial Transition Period

The framework establishes a graduated implementation schedule: a twelve-month preparation period, phased industry rollout, geographic staging, size-based implementation, and finally, pilot program testing. Key transition tools during this period will include legacy agreement management, grandfather provisions, temporary exemptions, compliance assistance, and technical support.

b. Market Adaptation

The system facilitates market adaptation in two ways. First, it provides industry-specific adjustment mechanisms, including sector-based timing, custom compliance tools, specialized guidance,

industry liaison offices, and targeted assistance programs. Second, it accounts for business size by offering accommodations for small businesses, support for medium enterprises, compliance programs for corporations, assistance for start-ups, and scalable resources.

2. Practical Implementation Solutions

a. Operational Challenges

The framework addresses key operational issues through concrete solutions.[247] Cost management will be achieved through automated compliance systems, shared service platforms, standardized documentation, electronic filing systems, and bulk processing options.

b. Compliance Assistance and Monitoring Solutions

To ensure effective implementation, the framework incorporates a comprehensive set of support mechanisms. Compliance assistance programs provide technical guidance centers, online compliance tools, training programs, advisory opinions, and help desk support. These resources focus on practical implementation through step-by-step compliance guides, interactive decision tools, template agreements, calculation worksheets, and best practices databases. Then, monitoring solutions incorporate real-time systems that feature automated compliance checks, early warning indicators, performance metrics, risk assessment tools, and pattern detection. In addition, data analytics play a role by leveraging predictive modeling, compliance trending, impact assessment, efficiency metrics, and outcome analysis.

3. Long-Term Sustainability

The framework ensures sustainable implementation through continuous improvement mechanisms by way of regular system evaluation, stakeholder feedback loops, adaptive management, performance optimization, and innovation incorporation. It also supports institutional learning through knowledge management systems, best practices evolution, precedent databases, training updates, and process refinement.

While the transition to a federal framework presents significant challenges, these obstacles can be overcome through careful planning and robust support systems. Success

247 *See* DAVID WEIL, IMPROVING WORKPLACE CONDITIONS THROUGH STRATEGIC ENFORCEMENT 1–2 (2010).

requires sustained commitment to implementation excellence and willingness to adapt based on empirical evidence and practical experience.

IX. CONCLUSION

The current debate over non-compete agreements in the United States has reached a critical juncture. The FTC's categorical ban, while addressing legitimate concerns about worker mobility, represents an overcorrection that fails to acknowledge the complex balance of interests at stake.[248] This Article has proposed an alternative approach: a federal framework based on the German model of compensated non-compete agreements.

The proposed framework offers several key advantages over both the status quo and the FTC's approach. First, mandatory compensation requirements create efficient market-based incentives that force employers to internalize restriction costs,[249] protect employee economic security, facilitate voluntary compliance, enable market-based adjustments, and preserve legitimate business interests. Second, the federal framework provides needed uniformity while maintaining appropriate flexibility through clear minimum standards, state implementation authority, industry-specific adaptations, size-based accommodations, and regional adjustments. Third, empirical evidence from Germany demonstrates that structured regulation produces superior outcomes in terms of innovation rates,[250] worker mobility, economic efficiency, litigation reduction, and knowledge protection.

The success of the German model provides compelling evidence that a compensatory approach can effectively balance competing interests while promoting economic dynamism. As the United States grapples with evolving workplace relationships and increasing competition for skilled workers, the need for sophisticated regulatory approaches becomes more pressing. The framework proposed in this Article offers a path forward that

---

248 *See* Brian Albrecht, *When Protection Becomes Overreach*, CITY J. (May 13, 2024), https://www.city-journal.org/article/the-ftc-ban-on-noncompete-agreements-is-misguided [https://perma.cc/NJ5M-HZYF].

249 BRIAN C. ALBRECHT, DIRK AUER & GEOFFREY A. MANNE, LABOR MONOPSONY AND ANTITRUST ENFORCEMENT: A CAUTIONARY TALE 30–31 (2024).

250 *See* Iain Ross, *Non-Compete Clauses in Employment Contracts: The Case for Regulatory Response*, 35 ECON. & LAB. RELS. REV. 806, 822 (2024).

learns from international experience while accounting for unique American legal and economic conditions.

Implementation challenges, while significant, are not insurmountable. The proposed framework's graduated implementation schedule, coupled with robust support systems and clear guidance, provides a realistic pathway to reform. Experience demonstrates that initial transition costs are typically offset by reduced litigation expenses and improved economic outcomes within three to five years of implementation.[251]

Moreover, the framework's flexibility allows it to adapt to emerging workplace trends. For example, the rise of remote work challenges traditional geographic restrictions, but these can be addressed through carefully crafted scope limitations and compensation adjustments. Similarly, the growing importance of knowledge-based competition calls for more sophisticated strategies to protect legitimate business interests while still encouraging innovation. Finally, the increasing mobility of skilled workers demands balanced approaches that protect both employer investments and employee career development.

Looking forward, the success of this framework will depend on sustained commitment from multiple stakeholders. First, Congress must provide clear statutory authority and adequate resources for implementation. Second, federal agencies must develop sophisticated enforcement mechanisms while coordinating effectively with state partners. Third, courts must adapt to new methods of analysis while maintaining consistent interpretation of standards. Finally, and above all else, employers and employees must engage constructively with the new system, recognizing that structured regulation can create value for all parties.

Ultimately, the choice is not between unfettered non-compete agreements and their complete prohibition, but rather between thoughtful regulation and continued uncertainty. The framework proposed here offers a tested, balanced approach that has succeeded in one of the world's most innovative economies.[252] As Congress considers various approaches to non-compete reform, the compensatory model deserves serious consideration as a solution

---

251 *See* Stam, *supra* note 170, at 6.

252 *See* Kate Whiting, *Germany Is the World's Most Innovative Economy*, WORLD ECON. F. (Oct. 18, 2018), https://www.weforum.org/stories/2018/10/germany-is-the-worlds-most-innovative-economy/ [https://perma.cc/BH62-HBB6] ("In the World Economic Forum's latest Global Competitiveness Report, Germany came top as the world's most innovative economy, with a score of 87.5 out of 100 in the Innovation capability pillar - one of the 12 drivers of a country's productivity.").

that serves both business interests and worker protection. By adopting this approach, the United States can move beyond the current regulatory impasse toward a system that promotes innovation, protects legitimate business interests, and ensures worker mobility and economic security. The time has come to replace our patchwork of state regulations and binary approaches with a sophisticated federal framework that reflects the complexities of modern employment relationships and the realities of our knowledge-based economy.

Social Media, Populism, and Militant Democracy: An Argument for a Change in First Amendment Doctrine

Arvid Kerschnitzki

Social Media, Populism, and Militant Democracy: An Argument for a Change in First Amendment Doctrine

Arvid Kerschnitzki[*]

The First Amendment is often hailed as the cornerstone of American self-government, but the digital age has transformed the public discourse it was meant to protect. Social media platforms, once envisioned as a modern marketplace of ideas, now amplify misinformation, entrench echo chambers, and enable mass manipulation, threatening the democratic ideal of an informed citizenry.

While the current First Amendment doctrine rightly emphasizes the importance of public discourse, that discourse must be inclusive, rational, and reliable to serve its democratic function. Drawing on the concept of militant democracy, this Article argues that contemporary doctrine has lost sight of free speech's dual purpose—not only to protect individual liberty, but also to safeguard the democratic process—and that it must be reoriented to reflect this foundational balance.

[*] Student and Research Assistant at the Junior Professorship for Public Law and at the Junior Professorship for Private Law at the University of Augsburg in Germany. Earned an LL.M. degree from The George Washington University Law School in 2024 (degree will be awarded after the graduation ceremony in Germany).

I. NEW CHALLENGES

The face of political propaganda has changed dramatically over the last twenty years. The internet, and social media in particular, have opened fundamentally new ways of influencing public opinion. Today's digital sphere is shaped by misinformation, rabble-rousing, and trolls.[1] This poses a tangible threat to democracy.[2] The link between social media and global democratic backsliding[3] is well established,[4] although its extent is up for debate.[5] This Article will discuss how these new

---

1 *See Lies, Bots, and Social Media: What Is Computational Propaganda and How Do We Defeat It?: Joint Hearing Before the Comm. on Sec. & Coop. in Eur.*, 115th Cong. 1–7 (2018) (statement of Mark Toner, State Department Senior Advisor, Commission on Security & Cooperation in Europe); *see also Rabble-Rousing*, COLLINS DICTIONARY, https://www.collinsdictionary.com/us/dictionary/english/rabble-rousing [https://perma.cc/2F4A-6YYM] (defining rabble-rousing as "encouragement that a person gives to a group of people to behave violently or aggressively, often for that person's own political advantage") (last visited Apr. 27, 2025); *Troll*, COLLINS DICTIONARY, https://www.collinsdictionary.com/us/dictionary/english/troll [https://perma.cc/RH4P-9ERB] (defining a troll as "a person who posts derogatory or abusive messages on the internet") (last visited Apr. 25, 2025).

2 *See* Newton Minow & Martha Minow, *Social Media Companies Should Pursue Serious Self-Supervision — Soon: Response to Professors Douek and Kadri*, 136 HARV. L. REV. F. 428, 438 (2023) ("Current and former social media executives warn that their industry harms civil discourse essential to democracy and amplifies misinformation, social divisions, and risks of violence.").

3 *See* Thomas Carothers & Benjamin Press, Understanding and Responding to Global Democratic Backsliding 4–6 (Oct. 20, 2022) (working paper) (on file with the Carnegie Endowment for International Peace), https://carnegie-production-assets.s3.amazonaws.com/static/files/Carothers_Press_Democratic_Backsliding_v3_1.pdf [https://perma.cc/33PM-SVZD]; *see also* RALPH SCHROEDER, SOCIAL THEORY AFTER THE INTERNET 60–81 (2018).

4 *See, e.g.*, Jonathan Haidt, *Yes, Social Media Really Is Undermining Democracy*, THE ATLANTIC (July 28, 2022), https://www.theatlantic.com/ideas/archive/2022/07/social-media-harm-facebook-meta-response/670975/ [https://perma.cc/M8C7-B4YK]; Guy Schleffer & Benjamin Miller, *The Political Effects of Social Media Platforms on Different Regime Types*, 4 TEX. NAT'L SEC. REV. 77, 78 (2021) (finding that social media weakens strong democratic regimes and radicalizes weak democratic regimes); Yphtach Lelkes et al., *The Hostile Audience: The Effect of Access to Broadband Internet on Partisan Affect*, 61 AM. J. POL. SCI. 5, 17 (2017) (concluding that "the new media environment has contributed to increased partisan animus and that greater exposure to biased news sources is the likely cause").

5 *See, e.g.*, Jan-Werner Müller, *The Myth of Social Media and Populism*, FOREIGN POL'Y (Jan. 3, 2024, 12:20 AM), https://foreignpolicy.com/2024/01/03/2024-elections-social-media-technology-democracy/ [https://perma.cc/9BXH-TDMS] (expressing skepticism with respect to the extent of social media's role in this crisis). The rise of populism is surely, above all, a problem of a shift in mentality, and not merely one of the law. *See, e.g.*, András Sajó, *Militant Constitutionalism*, *in* MILITANT DEMOCRACY AND ITS CRITICS: POPULISM, PARTIES, EXTREMISM 187, 187–89 (Anthoula Malkopoulou & Alexander S. Kirshner eds., 2019). However, this shift of mentality is heavily influenced by populist actors using sophisticated methods of propaganda and manipulation. Some claim that polarization today works in a top-down manner, meaning that polarization in the political sphere does not mirror the polarization of the society, but rather that political actors

technologies and methods of influencing the electorate via social media threaten democracy (Part I) and how these threats are not being addressed by the current First Amendment doctrine (Part II). Acknowledging that a shift in First Amendment doctrine is overdue, this Article will introduce the concept of militant democracy[6] (Section III.A) and conduct a textual and historical analysis of the U.S. Constitution to examine whether First Amendment doctrine is receptive to militant democracy-type arguments (Section III.B). At this point, it should be made clear that this Article does not claim—nor does it intend—to solve all of the problems of First Amendment doctrine and social media. Rather, it simply advocates for the use of the argumentative figure of militant democracy in the First Amendment debate.

A. New Technology, New Actors, New Methods

Social media has not only drastically changed private communication, but it has also become a major source of news for a large part of the population.[7] The public exchange of political opinions has also largely shifted from analog forums to social media platforms.[8] This paradigm shift has long been seen as a huge democracy booster, enabling unprecedented levels of

actively work towards the latter for their own political profit. *See, e.g.*, YOCHAI BENKLER ET AL., NETWORK PROPAGANDA 305 (2018); *Political Polarization*, EUR. CTR. FOR POPULISM STUD., https://www.populismstudies.org/vocabulary/political-polarization/ [https://perma.cc/K4WQ-E4XY] (last visited Apr. 4, 2025); Hubert Tworzecki, *Poland: A Case of Top-Down Polarization*, 681 ANNALS AM. ACAD. POL. & SOC. SCI. 97, 100 (2019); STEFFEN MAU ET AL., TRIGGERPUNKTE: KONSENS UND KONFLIKT IN DER GEGENWARTSGESELLSCHAFT 322–33 (2023) (discussing this phenomenon in Germany and the United States). *But see* Daniel J. Hopkins et al., *From Many Divides, One? The Polarization and Nationalization of American State Party Platforms, 1918–2017*, 36 STUD. AM. POL. DEV. 1, 20 (2022) (finding that the evidence of polarization in fact "undercut[s] claims that [the] phenomenon is principally a top-down process"). Still, populists' main medium for such manipulation is social media. So, the question of how to combat propaganda and manipulation via social media goes hand in hand with the question of how to combat this shift in mentality.

6 This is sometimes also called "Defensive Democracy." *See* Jon Smibert, *Defensive Rule of Law*, 73 EMORY L.J. 1111, 1139 (2024).

7 *See, e.g.*, Nic Newman, *Overview and Key Findings of the 2023 Digital News Report*, REUTERS INST. (June 14, 2023), https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2023/dnr-executive-summary [https://perma.cc/TJX6-29EC]; *Social Media and News Fact Sheet*, PEW RSCH. CTR. (Sept. 17, 2024), https://www.pewresearch.org/journalism/fact-sheet/social-media-and-news-fact-sheet/ [https://perma.cc/9QJ3-BRT9]; Andrew Hutchinson, *New Research Shows that 71% of Americans Now Get News Content via Social Platforms*, SOC. MEDIA TODAY (Jan. 12, 2021), https://www.socialmediatoday.com/news/new-research-shows-that-71-of-americans-now-get-news-content-via-social-pl/593255/ [https://perma.cc/P2DQ-VPJY].

8 *See* Packingham v. North Carolina, 582 U.S. 98, 99 (2017) (noting that social media is "the modern public square").

democratic participation.[9] But the very nature of social media has turned it into the opposite of what it promised: it has become a democratic Pandora's box.

As with all forms of communication, social media is used by a wide variety of actors to influence people.[10] Before social media, these actors were mainly domestic politicians or, in some instances, foreign state actors. The landscape of public opinion was comparatively small. Social media opened up a whole new public sphere for anyone with access to the Internet to participate in the public debate. This was undoubtedly a victory for democratic participation.[11]

However, it soon became apparent that this space offered an unprecedented potential for influencing public opinion. Among the plethora of new actors, foreign states use social media to spread propaganda[12] and influence elections,[13] often through the use of bots. Foreign terrorist groups use social media to recruit followers and to incite terrorist attacks.[14] Additionally, certain private actors—such as conspiracy theorists—suddenly have an audience of previously unknown size.[15]

---

9 Despite some evidence to the contrary, the fact that social media promotes participation remains true. *See* Joshua A. Tucker et al., *From Liberation to Turmoil: Social Media and Democracy*, 28 J. DEMOCRACY 46, 48–50 (2017) (noting that social media provides "new and expressive forms for participation in the political process," including "mini-participation").

10 *See* Nicola A. Boothe-Perry, *Friends of Justice: Does Social Media Impact the Public Perception of the Justice System?*, 35 PACE L. REV. 72, 90 (2014) (discussing a Florida law that calls for automatic disqualification of a judge when they are friends on Facebook with a lawyer who is appearing before them, due to concerns of improper influence); *see also* NAT'L INTEL. COUNCIL, OFF. OF THE DIR. OF NAT'L INTEL., Assessing Russian Activities and Intentions in Recent U.S. Elections, at ii (2017) ("Moscow's influence campaign followed a Russian messaging strategy that blends covert intelligence operations—such as cyber activity—with overt efforts by Russian Government agencies, state-funded media, third-party intermediaries, and paid social media users or 'trolls.'").

11 *But see* Karen Kornbluh, *The Internet's Lost Promise: And How America Can Restore It*, 97 FOREIGN AFFS. 33, 34–35 (2018) (finding that hope in the internet aiding democracy was misguided).

12 This includes the sophisticated deployment of paid influencers to spread certain political messages, even with memes. *See* Samuel C. Woolley, *Digital Propaganda: The Power of Influencers*, 33 J. DEMOCRACY 115, 120 (2022).

13 *See* Kornbluh, *supra* note 11, at 36–37.

14 *See* Zachary Leibowitz, *Terror on Your Timeline: Criminalizing Terrorist Incitement on Social Media Through Doctrinal Shift*, 86 FORDHAM L. REV. 795, 810 (2017).

15 *See* Matteo Cinelli et al., *Conspiracy Theories and Social Media Platforms*, 47 CURRENT OP. PSYCH., Oct. 2022, at 1, 1, 4 (finding that "[c]onspiracy theories proliferate online" and attributing this phenomenon to increased polarization and echo chambers, but also possibly to "the role of recommendation algorithms and moderation policies").

Furthermore, as privately owned companies, social media platforms are designed to maximize profits. They do this by promoting attention-grabbing posts and automatically pushing emotionally inflammatory content,[16] since this type of content achieves the most user interactions.[17] Such emotional, fast-spreading content often includes misinformation, hate speech, and rabble-rousing, which have been shown to generate extremely high levels of user engagement.[18] To increase user engagement, social media platforms' algorithms also tend to provide users with posts that reflect their own points of view, creating so-called "echo chambers."[19]

Moreover, it turned out that the assumed equality of speakers in the virtual realm was a deception. Money could buy influence not only in traditional media, but also online. The new technologies opened up the possibility of influencing mass opinion in a much more subtle way.[20] The use of bots, for example, allows actors to create the impression of widespread support for an opinion and significantly increase the reach of a message.[21] Spreading targeted dis- and misinformation,[22] or just

---

16 *See* SAMUEL WOOLLEY, MANUFACTURING CONSENSUS: UNDERSTANDING PROPAGANDA IN THE ERA OF AUTOMATION AND ANONYMITY 120–22 (2023); *see also* Steve Rathje et al., *Out-Group Animosity Drives Engagement on Social Media*, PROC. NAT'L ACAD. SCIS., June 23, 2021, at 1, 1, 7 (discussing the effect of online misinformation on political polarization).

17 This also explains why the social media posts of right-wing parties usually achieve the highest user engagement. *See* JUAN CARLOS MEDINA SERRANO ET AL., SOCIAL MEDIA REPORT: THE 2017 GERMAN FEDERAL ELECTIONS 36–37 (2018).

18 *See, e.g.*, Steven Lee Myers, *How Social Media Amplifies Misinformation More than Information*, N.Y. TIMES (Oct. 13, 2022), https://www.nytimes.com/2022/10/13/technology/misinformation-integrity-institute-report.html [https://perma.cc/B9DZ-SHHA]; Andrew Hutchinson, *New Study Shows that Misinformation Sees Significantly More Engagement than Real News on Facebook*, SOC. MEDIA TODAY (May 22, 2019), https://www.socialmediatoday.com/news/new-study-shows-that-misinformation-sees-significantly-more-engagement-than/555286/ [https://perma.cc/W5WY-E7CW]; Gilad Edelman, *Fake News Gets More Engagement on Facebook—But Only if It's Right-Wing*, WIRED (Mar. 3, 2021, 8:00 AM), https://www.wired.com/story/right-wing-fake-news-more-engagement-facebook/ [https://perma.cc/2LDF-Q2FR].

19 *See* Matteo Cinelli et al., *The Echo Chamber Effect on Social Media*, PROC. NAT'L ACAD. SCI., Mar. 2, 2021, at 1, 1–2, 5; Ludovic Terren & Rosa Borge, *Echo Chambers on Social Media: A Systematic Review of the Literature*, 9 REV. COMMC'N RSCH. 99, 100, 108–11 (2021).

20 For a broad overview of virtual manipulation methods, see WOOLLEY, *supra* note 16, at 22–28.

21 *See* Tim Wu, *Is the First Amendment Obsolete?*, 117 MICH. L. REV. 547, 548, 567 (2018); *see also* Robin Graber & Thomas Lindemann, *Neue Propaganda im Internet. Social Bots und das Prinzip Sozialer Bewährtheit als Instrumente der Propaganda*, *in* FAKE NEWS, HASHTAGS & SOCIAL BOTS 51, 57 (2018).

flooding the market with it,[23] became an easy and popular tool to confuse the public debate. Due to its sheer amount (and due to echo chambers), repeated exposure to false information increases the likelihood that users will perceive it as true.[24] Moreover, the aforementioned methods of hate speech and rabble-rousing have proven to be effective tools for emotionalizing public opinion and public discourse.[25]

B. New Instability

History has shown that democracy is an inherently fragile system.[26] Social media amplifies this instability. Democracy builds on "broadly accepted, legitimate political authority, some basic consensus regarding how to distinguish truth from falsity, and a sense that even ardent political opponents are part of the same polity, bound by a common fate."[27] Social media platforms "radically undermine those pillars."[28] Authoritarian propaganda and ethno-nationalist extremism have become a permanent and defining part of social media.[29] The characteristics mentioned above and the functioning of social media, such as its tendency to create echo chambers, have a broad effect on driving polarization.[30] Due to social media, "groups of like-minded

---

22 *See* Neil Netanel, *Applying Militant Democracy to Defend Against Social Media Harms*, 45 CARDOZO L. REV. 489, 507 (2023). The difference between misinformation and disinformation is often described in the way that misinformation is simply false or inaccurate, whereas disinformation deliberately intends to mislead. *Misinformation and Disinformation*, AM. PSYCH. ASS'N, https://www.apa.org/topics/journalism-facts/misinformation-disinformation [https://perma.cc/5YNM-UYK8] (last visited Apr. 10, 2025). This Article will (with some hesitation) stick with the term "misinformation."

23 *See* Anya Schiffrin, *Disinformation and Democracy: The Internet Transformed Protest but Did Not Improve Democracy*, 71 J. INT'L AFFS. 117, 118 (2017); *see also* Sean Illing, *"Flood the Zone with Shit": How Misinformation Overwhelmed Our Democracy*, VOX (Feb. 6, 2020, 6:27 AM), https://www.vox.com/policy-and-politics/2020/1/16/20991816/impeachment-trial-trump-bannon-misinformation [https://perma.cc/H5PF-8G3Y].

24 *See* Lisa K. Fazio et al., *Repetition Increases Perceived Truth Equally for Plausible and Implausible Statements*, 26 PSYCHONOMIC BULL. & REV. 1705, 1709–10, 1709 fig.2 (2019).

25 *See* Netanel, *supra* note 22, at 504.

26 *See* Sajó, *supra* note 5, at 187 ("Democracy is one of the gravest threats to democracy.").

27 Netanel, *supra* note 22, at 492 (citing Robert Post, *The Unfortunate Consequences of a Misguided Free Speech Principle*, 153 DÆDALUS 135, 143 (2024)).

28 *Id.*

29 *See id.* at 495.

30 Cass Sunstein calls this phenomenon "cyberpolarization." CASS R. SUNSTEIN, #REPUBLIC: DIVIDED DEMOCRACY IN THE AGE OF SOCIAL MEDIA 68 (rev. ed. 2018). Sunstein acknowledges, however, that traditional media has driven polarization before (even though not to such an extent). *See id.* at 61–62 (citing Gregory J. Martin & Ali Yurukoglu, *Bias in Cable News: Persuasion and Polarization*, 107 AM. ECON. REV. 2565,

people . . . will typically end up thinking the same thing that they thought before—but in a more extreme form."[31] Along with increasing polarization, social media intensifies the emotionalization and de-rationalization of public discourse.[32] Social media's heavy reliance on emotions creates an environment that leads to irrationally motivated political decisions.[33]

Due to the flood of misinformation, it has become considerably more difficult for individual citizens to distinguish real from fabricated news.[34] Among this plethora of misinformation, users tend to select information that fits their pre-existing beliefs, thereby reinforcing those very beliefs—as they do in the analogous world.[35] The effects of misinformation on democracy are manifold, ranging from basic (at least attempted) manipulation of voters,[36] to spreading distrust in

---

2565 (2017) (examining how MSNBC and Fox News correlate with and influence polarization across the American public)); *see also* Netanel, *supra* note 22, at 504.

31 SUNSTEIN, *supra* note 30, at 68, 76–79.

32 *See* Brian E. Weeks & R. Kelly Garrett, *Emotional Characteristics of Social Media and Political Misperceptions*, *in* JOURNALISM AND TRUTH IN AN AGE OF SOCIAL MEDIA 236, 238–42 (James E. Katz & Kate K. Mays eds., 2019).

33 *See* Philipp Lorenz-Spreen et al., *A Systematic Review of Worldwide Causal and Correlational Evidence on Digital Media and Democracy*, 7 NATURE HUM. BEHAV. 74, 85 (2023); *see also* Cameron Martel et al., *Reliance on Emotion Promotes Belief in Fake News*, COGNITIVE RSCH.: PRINCIPLES & IMPLICATIONS 15–20 (Oct. 7, 2020), https://cognitiveresearchjournal.springeropen.com/counter/pdf/10.1186/s41235-020-00252-3.pdf [https://perma.cc/3JD5-49VC] (exploring the connection between emotionalism and fake news); EVA ILLOUZ, THE EMOTIONAL LIFE OF POPULISM 162 (2023) (discussing the role of emotions in populist propaganda and identifying four main emotions that are exploited by right-wing populists).

34 *See* Philip M. Napoli, *What if More Speech Is No Longer the Solution? First Amendment Theory Meets Fake News and the Filter Bubble*, 70 FED. COMMC'NS L.J. 55, 79–85 (2018).

35 *See* Nicole M. Krause et al., *Fake News: A New Obsession with an Old Phenomenon?*, *in* JOURNALISM AND TRUTH IN AN AGE OF SOCIAL MEDIA 58, 66–68 (2019).

36 *See, e.g.*, Napoli, *supra* note 34, at 93–97 (discussing the role of fake news and echo chambers in the 2016 election); Hunt Allcott & Matthew Gentzkow, *Social Media and Fake News in the 2016 Election*, 31 J. ECON. PERSPS. 211, 211–12 (2017) (engaging in extensive discussions on the amount of misinformation in the election); Abigail Abrams, *Here's What We Know So Far About Russia's 2016 Meddling*, TIME (Apr. 18, 2019, 8:20 AM), https://time.com/5565991/russia-influence-2016-election/ [https://perma.cc/FJ5N-HQ4N] (considering Russia's involvement in this context); Brandy Zadrozny, *Disinformation Poses an Unprecedented Threat in 2024—and the U.S. Is Less Ready than Ever*, NBC NEWS (Jan. 18, 2024, 1:20 PM), https://www.nbcnews.com/tech/misinformation/disinformation-unprecedented-threat-2024-election-rcna134290 [https://perma.cc/M5AC-7SM6] (contemplating the role of disinformation in the then-upcoming 2024 presidential election); Tiffany Hsu et al., *Elections and Disinformation Are Colliding like Never Before in 2024*, N.Y. TIMES, https://www.nytimes.com/2024/01/09/business/media/election-disinformation-2024.html [https://perma.cc/5EU7-26BL] (Jan. 11, 2024) (discussing this issue on a global scale).

democratic institutions[37] and eroding the public's confidence in democracy,[38] to election denialism[39] and playing a major role in the run-up to the January 6, 2021 storm on the Capitol.[40] Of course, not all social media phenomena that influence elections are problematic. However, a democracy must be aware of the power of social media to sway elections[41] and how that power can be exploited—especially through misinformation.[42]

In particular, populist parties (usually right-leaning) have long since realized the power of social media and begun to use it to their advantage. Due to the emotional content of their campaigns, which translates into their online activities, those parties are generally more successful on social media.[43] In addition, the use of

---

37 *See* Carme Colomina et al., *The Impact of Disinformation on Democratic Processes and Human Rights in the World,* at 13–14, PE 653.635 (Apr. 2021), https://www.europarl.europa.eu/RegData/etudes/STUD/2021/653635/EXPO_STU(2021)653635_EN.pdf [https://perma.cc/6356-VLPV].

38 *See* Gabriel R. Sanchez & Keesha Middlemass, *Misinformation Is Eroding the Public's Confidence in Democracy*, BROOKINGS (July 26, 2022), https://www.brookings.edu/articles/misinformation-is-eroding-the-publics-confidence-in-democracy/ [https://perma.cc/6BEX-2WAJ].

39 PAUL M. BARRETT, N.Y.U. STERN CTR. FOR BUS. & HUM. RTS., SPREADING THE BIG LIE: HOW SOCIAL MEDIA SITES HAVE AMPLIFIED FALSE CLAIMS OF U.S. ELECTION FRAUD 1 (2022), https://bhr.stern.nyu.edu/tech-big-lie [https://perma.cc/W3NP-7M4G].

40 *See* SELECT COMM. TO INVESTIGATE THE JANUARY 6TH ATTACK ON THE U.S. CAPITOL, FINAL REPORT OF THE SELECT COMMITTEE TO INVESTIGATE THE JANUARY 6TH ATTACK ON THE UNITED STATES CAPITOL, H.R. REP. NO. 117-663, at 213 (2022), https://www.congress.gov/117/crpt/hrpt663/CRPT-117hrpt663.pdf [https://perma.cc/LY79-HLYP]; *see also* Wes Henricksen, *Disinformation and the First Amendment: Fraud on the Public*, 96 ST. JOHN'S L. REV. 543, 546–47 (2022) (discussing the connection between the "Big Lie" and the storm on the Capitol).

41 This applies not only to long-term influence over social media, but also to short-term disruptions via social media. For example, YouTuber Rezo released a YouTube video shortly before the German federal elections, titled *Die Zerstörung der CDU* [*The Destruction of the CDU*], whereafter the Christian Democratic Union—the center-conservative party—plunged seven percent in the polls. *See* Rezo, *Die Zerstörung der CDU.*, YOUTUBE (May 18, 2019), https://www.youtube.com/watch?v=4Y1lZQsyuSQ [https://perma.cc/NC4Z-RJLP]; *see also* Richard H. Pildes, *Democracies in the Age of Fragmentation*, 110 CAL. L. REV. 2051, 2059–60 (2022); Peter Kuras, *German Politics Discovers YouTube*, FOREIGN POL'Y (June 4, 2019, 5:23 PM), https://foreignpolicy.com/2019/06/04/german-politics-discovers-youtube/ [https://perma.cc/63EK-JKFS].

42 *See* Henricksen, *supra* note 40, at 554 (analyzing more tangible harms caused by disinformation).

43 *See Why Right-Wing Populist Parties Have a Greater Reach*, KLICKSAFE (Mar. 11, 2024), https://www.klicksafe.de/en/news/warum-rechtspopulisten-reichweitenstaerker-sind [https://perma.cc/9MW9-9PBC]; *see also* Sandra González-Bailón et al., *The Advantage of the Right in Social Media News Sharing*, PNAS NEXUS, July 29, 2022, at 1, 1 (discussing the "asymmetries in the ideological slant . . . on social media, with a clear bias towards right-leaning domains"); SCHROEDER, *supra* note 3, at 79–81 (discussing populism and social media in the United States, Sweden, India, and China); Madelaine Pitt & Hans Pfeifer, *Far-Right AfD Is a Social Media Superpower*, DEUTSCHE WELLE

bots by right-wing and anti-establishment parties makes their supporters feel like they are part of a larger movement.[44] This sense of community normalizes previously morally or socially discredited language or ideas in the public discourse.[45]

Adding hate speech to that equation, which heightens the emotionality even further, social media seems to be, at least in large areas, primarily composed of "uncivil, manipulative free-for-all zones."[46] The mechanisms in place and the potential for abuse pose serious threats to democracy.[47] Many aspects of social media as it operates today fundamentally undermine democratic coexistence[48] and seem to be one of the reasons for the rising distrust in democracy and democratic institutions around the world.[49]

(Aug. 27, 2021), https://www.dw.com/en/german-election-far-right-afd-outperforms-competitors-on-social-media/a-59004003 [https://perma.cc/K9GW-T9SH] (referring to the Alternative for Germany political party, or AfD, as a "social media superpower"). One reason for its popularity is certainly the considerable extra effort that populist parties put into their social media campaigns.

44 *See* Juan Carlos Medina Serrano et al., *The Rise of Germany's AfD: A Social Media Analysis*, *in* SMSOCIETY '19: PROCEEDINGS OF THE 10TH INTERNATIONAL CONFERENCE ON SOCIAL MEDIA AND SOCIETY 214, 216, 222 (2019) (finding that social media has become the AfD's main communication tool); *see also* JOHANNES HILLJE, DAS »WIR« DER AFD: KOMMUNIKATION UND KOLLEKTIVE IDENTITÄT IM RECHTSPOPULISMUS 64–65, 232 (2022) (noting that the AfD uses its social media posts to construct a collective identity, and identifying the characteristic elements of populism—specifically abbreviation, polarization, and emotionalization—as virality factors for social media content).

45 *See* Mats Ekström et al., *The Normalization of the Populist Radical Right in News Interviews: A Study of Journalistic Reporting on the Swedish Democrats*, 30 SOC. SEMIOTICS 466, 466–68 (2020). In this context, Germans use the term of making this language "salonfähig" (meaning "socially acceptable"). James Angelos, *Germany's Far-Right 'Firewall' Cracks*, POLITICO (Oct. 4, 2023, 4:00 AM), https://www.politico.eu/article/germany-firewall-afd-elections-thuringia/ [https://perma.cc/8RH9-AYR9].

46 Netanel, *supra* note 22, at 500.

47 *See, e.g.*, Lorenz-Spreen et al., *supra* note 33, at 83; Jaeho Cho et al., *Do Search Algorithms Endanger Democracy? An Experimental Investigation of Algorithm Effects on Political Polarization*, 64 J. BROAD. & ELEC. MEDIA 150, 166–68 (2020) (explaining the effects of algorithms and polarization); Antoine Banks et al., *#PolarizedFeeds: Three Experiments on Polarization, Framing, and Social Media*, 26 INT'L J. PRESS/POL. 609, 630–31 (2021).

48 For an extensive discussion on how social media destabilizes democracies, see Netanel, *supra* note 22, at 498. Later in his article, Netanel emphasizes that "social media undermine[s] the fundamental epistemic predicate of pluralist democracy." *Id.* at 507. Additionally, there are other ways in which the First Amendment undermines democracy. *See* CASS R. SUNSTEIN, DEMOCRACY AND THE PROBLEM OF FREE SPEECH 119 (rev. ed. 1995). Sunstein already feared the dangers that would come with new technologies even before the rise of social media. *Id.* at 257–58.

49 *See* Schleffer & Miller, *supra* note 4, at 85–86.

II. THE INSUFFICIENCY OF THE CURRENT DOCTRINE

In the United States, these threats are only half-heartedly addressed, if at all. There are several reasons for this failure—among the most important is the American technology-liberal and techno-utopian policy approach, which emphasizes minimal regulation of digital technologies and a strong belief in technological innovation as an inherently positive force for societal progress.[50] Yet, even if the political landscape and the policy-making approach were to change, current First Amendment doctrine will remain an insurmountable barrier to most policy changes.[51]

A. Origins: Holmes and Brandeis

To understand the current First Amendment doctrine, one must first understand the purpose of the Free Speech Clause of the U.S. Constitution. The purpose of freedom of speech in a liberal democracy is twofold. On the one hand, the right to free speech is an end—an expression that, by its mere utterance, reaffirms the Constitution's guarantee of individual liberty.[52] Yet free speech is also fundamentally a means—an essential step toward participating in democratic self-government. This dual function has been illuminated in particular by the perspectives of Supreme Court Justices Oliver Wendell Holmes, Jr., and Louis D. Brandeis.[53]

Justice Brandeis believed that the First Amendment is "essential to effective democracy" and could be limited only when "the evil apprehended is relatively serious."[54] His core belief was that the First Amendment serves democracy[55] because it enables the conditions necessary for democracy, such as an informed

---

50 *See* Netanel, *supra* note 22, at 520, 525; *see also* Paul Starr, *How Neoliberal Policy Shaped the Internet—and What to Do About It Now*, AM. PROSPECT (Oct. 2, 2019), https://prospect.org/api/content/e4fb9f4a-e460-11e9-9156-12f1225286c6/ [https://perma.cc/Z9F2-ZEL5] (discussing the history of neoliberal internet policy).

51 *See* Brittany Finnegan, Note, *The Cost of Free Speech: Combating Fake News or Upholding the First Amendment?*, 75 U. MIA. L. REV. 572, 618–19 (2021).

52 For example, James Madison called the freedom of the press (and implicitly of speech) "one of the great bulwarks of liberty" in his version of the Free Speech and Press Clauses, which were defeated in the Senate. 1 ANNALS OF CONG. 451 (1789) (Joseph Gales ed., 1834).

53 *See* Netanel, *supra* note 22, at 529.

54 Whitney v. California, 274 U.S. 357, 377 (1927) (Brandeis, J., concurring); *see also* David M. Rabban, *The Emergence of Modern First Amendment Doctrine*, 50 U. CHI. L. REV. 1205, 1340 (1983).

55 For a discussion of Justice Brandeis' optimism in this regard, see PHILIPPA STRUM, BRANDEIS ON DEMOCRACY 208–10 (1995).

electorate and a battle of ideas. His overall optimistic view was based on his strong belief in counterspeech. He was convinced that counterspeech is an appropriate and effective remedy for harmful speech, and that this desirable battle of ideas would ultimately strengthen both "self-government and individual liberty."[56]

Less optimistic than Justice Brandeis was his counterpart, Justice Holmes. In the spirit of the "marketplace of ideas" approach, which is often attributed to him,[57] Holmes emphasized the individual liberty function of free speech rather than its role in democracy. He seemed convinced that freedom of speech is an end in itself, perhaps even a liberty that is more important than democracy itself. His opinion in *Gitlow v. New York* supports this claim: "If in the long run the beliefs expressed in proletarian dictatorship are destined to be accepted by the dominant forces of the community, the only meaning of free speech is that they should be given their chance and have their way."[58] This quotation demonstrates how Justice Holmes was generally not so much concerned with the tyranny of the majority,[59] but rather he believed that freedom of speech was primarily an objective, meant to promote individual liberty[60] and pose a balancing "counterweight . . . to illiberal attitudes."[61]

Justice Holmes' strong belief in the "free trade in ideas"[62] came to shape the ultra-liberal First Amendment doctrine. Subsequent First Amendment jurisprudence, however, makes clear that the First Amendment and democracy are inextricably

---

56 Netanel, *supra* note 22, at 529 (citing *Whitney*, 274 U.S. at 377 (Brandeis, J., concurring)).

57 This theory was first mentioned as "competition of the market" in *Abrams v. United States*, 250 U.S. 616, 630 (1919) (Holmes, J., dissenting). For insight into Justice Holmes' understanding, see Vincent Blasi, *Holmes and the Marketplace of Ideas*, 2004 SUP. CT. REV. 1, 2 (describing Holmes' understanding as having "more to do with checking, character, and culture than with the implausible vision of a self-correcting, knowledge-maximizing, judgment-optimizing, consent-generating, and participation-enabling social mechanism").

58 Gitlow v. New York, 268 U.S. 652, 673 (1925) (Holmes, J., dissenting).

59 *See* Blasi, *supra* note 57, at 30–31. Interestingly, Justice Holmes' philosophy differs greatly from that of John Stuart Mill. Mill, who is often seen as the precursor to the "marketplace of ideas" doctrine, feared that democracy would become a tyranny of the majority and thus lead to an oppression purportedly legitimated by democratic popular will. *See* JOHN STUART MILL, ON LIBERTY 10–11, 18–19 (Floating Press 2009) (1909).

60 Meiklejohn later criticized Justice Holmes' philosophy for being "one of excessive individualism." ALEXANDER MEIKLEJOHN, POLITICAL FREEDOM: THE CONSTITUTIONAL POWERS OF THE PEOPLE 61 (1965).

61 Blasi, *supra* note 57, at 46.

62 For a discussion on the origin of this approach, see Dawn Carla Nunziato, *The Marketplace of Ideas Online*, 94 NOTRE DAME L. REV. 1519, 1523–27 (2019).

linked,[63] and that the First Amendment is a "guardian of our democracy."[64] Illustrative here is the work of American philosopher Alexander Meiklejohn, who, in many ways clearly opposed Justice Holmes' understanding of the First Amendment.[65] He strongly believed that the First Amendment is not an end in and of itself.[66] Instead, Meiklejohn was convinced that the "most significant purpose of the First Amendment" is to serve as the "intellectual basis of our plan of self-government."[67] With this in mind, it was clear to him that freedom of speech is absolute only when it pertains to the public interest.[68] In his view, Justice Holmes' "competitive individualism" ultimately "robs the [A]mendment of its essential meaning—the meaning of our common agreement that . . . we will be our own rulers."[69]

What has prevailed over all these philosophical disputes is the understanding that the First Amendment plays a vital role in American democracy. This may not be its sole purpose, but it is one of utmost importance.[70] Essential to that understanding has always been a strong belief in counterspeech: the belief that speech regulates itself. Whether one views counterspeech as a tool to arrive at "truth" in social discourse or as an alternative to government regulation,[71] it is fundamental to the current understanding of free speech.

B. The Current Doctrine

Restrictions of speech have traditionally been permitted only in limited circumstances. And despite the wide variety of First Amendment doctrines, these restrictions all seem to have one

---

63 *See* SUNSTEIN, *supra* note 30, at 212 (arguing that "the First Amendment in large part embodies a democratic ideal").

64 Brown v. Hartlage, 456 U.S. 45, 60 (1982).

65 As discussed above, Meiklejohn did not tire of pointing out the "failure of Mr. Holmes to recognize the sane and solid moral principles which find expression in our national agreement that government shall be carried on only by consent of the governed." MEIKLEJOHN, *supra* note 60, at 68.

66 Meiklejohn was of the opinion that the "First Amendment . . . is not the guardian of unregulated talkativeness. . . . What is essential is . . . that everything worth saying shall be said." *Id.* at 26.

67 *Id.* at 30.

68 *Id.* at 39. The public interest in speech sits in contradistinction to the private interest in speech, which is not essential to self-governance and thus may be abridged. *Id.* at 55.

69 *Id.* at 75.

70 *See* SUNSTEIN, *supra* note 30, at 212; *see also* OWEN M. FISS, THE IRONY OF FREE SPEECH 2–4 (1996).

71 G.S. Hans, *Changing Counterspeech*, 69 CLEVELAND STATE L. REV. 749, 774 (2021).

thing in common: the speech to be restricted must be expected to create some sort of imminent harm.[72] In the words of Justice Holmes, "[t]he question in every case is whether the words used are used in such circumstances and are of such a nature as to create a clear and present danger that they will bring about the substantive evils that Congress has a right to prevent."[73] This clear and present danger test was later incorporated into the incitement doctrine established in *Brandenburg v. Ohio*.[74]

According to this philosophy, First Amendment doctrine focuses on the fact that speech can only be restricted if the potential danger materializes very soon.[75] All other speech, no matter how repugnant, is considered awful but lawful.

Furthermore, based on case law, the First Amendment does not seem to require any kind of fairness in the public sphere. Although the Supreme Court ruled in *Red Lion Broadcasting Co. v. FCC* that the Federal Communications Commission's "fairness doctrine" was constitutional because it enhanced each individual's freedom of speech,[76] it did not hold in that case or declare in any subsequent jurisprudence that such fairness is generally required by the First Amendment. The Court's main holding in *Red Lion* was that radio frequencies are a limited resource and therefore must be fairly allocated.[77] This could lead to the conclusion that the *Red Lion* doctrine does not apply to social media because social media eliminates the scarcity of communication channels.[78] The Supreme Court has also explicitly stated that the First Amendment does not recognize

---

72 *See* Brandenburg v. Ohio, 395 U.S. 444, 452–53 (1969) (incitement); Chaplinsky v. New Hampshire, 315 U.S. 568, 572–73 (1942) (fighting words); N.Y. Times Co. v. Sullivan, 376 U.S. 254, 300 (1964) (false statements of facts); Ginsberg v. New York, 390 U.S. 629, 641 (1968) (obscenity regarding minors).

73 Schenck v. United States, 249 U.S. 47, 52 (1919).

74 *See generally* Chester James Antieau, *Clear and Present Danger—Its Meaning and Significance*, 25 NOTRE DAME L. REV. 603, 643–45 (1950) (displaying uncertainty about the subjectivity of this test). *See also* MEIKLEJOHN, *supra* note 60, at 47–49 (criticizing the test for its lack of objectivity).

75 Restriction is also permitted where the infliction of harm is inherent in the production of this sort of speech. *See* New York v. Ferber, 458 U.S. 747, 747–48, 757–59 (1982) (illustrating the doctrine as applied to child sexual abuse material). In fact, in *Ashcroft*, the Court held that the government may not prohibit speech because it increases the chance that an unlawful act will be committed at some indefinite time in the future. Ashcroft v. Free Speech Coal., 535 U.S. 234, 253–54 (2002).

76 Red Lion Broad. Co. v. FCC, 395 U.S. 367, 375 (1969).

77 *Id.* at 383, 388–89, 394.

78 *See* Wu, *supra* note 21, at 577–78 (believing that the holding of *Red Lion* is inapplicable and that imposing a fairness doctrine on social media would be constitutional).

political equality or the erosion of public debate as grounds for restricting speech.[79] In its 2024 case, *Moody v. Netchoice, LLC,* the Supreme Court indicated its approval of content moderation as part of the platform's editorial choice and thereby strengthened the editorial rights of the speech distributors.[80] Again, the fundamental right perspective prevails.

In addition, social media has been compared to a public forum, which leads to even higher standards of protection.[81] Concerning the threatened TikTok ban,[82] it remains to be seen whether the denial of complete access to a platform for everyone destroys an established public forum, and whether it will be considered consistent with the First Amendment.[83]

C. The Misfit

This approach to the First Amendment is inadequate to deal with the new challenges posed by social media.[84] There are huge differences between traditional mass media and social media, the latter of which poses unique threats to democracy.[85] A "business-as-usual" approach does not work.[86]

The premise of the current First Amendment doctrine, a self-regulating marketplace of ideas, is out of balance.[87] Due to new

---

79 *See* Citizens United v. FEC, 558 U.S. 310, 349–50 (2010).

80 Moody v. Netchoice, LLC, 603 U.S. 707, 708, 710–11 (2024).

81 The public forum doctrine is often traced back to Justice Roberts' opinion in *Hague v. Committee for Industrial Organization*, 307 U.S. 496, 515–17 (1939).

82 *See* Bobby Allyn, *Trump Extends TikTok's Sell-by Deadline Again*, NPR (Apr. 4, 2025, 2:07 PM), https://www.npr.org/2025/04/04/nx-s1-5347418/trump-tiktok-second-ban-delay [https://perma.cc/Q53M-96RC].

83 The American Civil Liberties Union considers this bill to be in violation of the First Amendment. *See ACLU Slams House for Latest Plan to Ban TikTok and Stifle Free Speech*, ACLU (Mar. 5, 2024, 6:00 PM), https://www.aclu.org/press-releases/aclu-slams-house-for-latest-plan-to-ban-tiktok-and-stifle-free-speech [https://perma.cc/MD34-BE3Y]. *See generally* Scott Bomboy, *A National TikTok Ban and the First Amendment*, NAT'L CONST. CTR. (Mar. 22, 2024), https://constitutioncenter.org/blog/a-national-tiktok-ban-and-the-first-amendment [https://perma.cc/5NCP-F8U5] (describing the controversies in lower courts).

84 *See* Nunziato, *supra* note 62, at 1527 (discussing the unique problems social media poses to the marketplace of ideas doctrine).

85 *See* Robert C. Post, *Democracy and the Internet*, BALKINIZATION (Jan. 28, 2023, 9:30 AM), https://balkin.blogspot.com/2023/01/democracy-and-internet.html [https://perma.cc/MW7S-KYLL].

86 *See* Lee C. Bollinger & Geoffrey R. Stone, *Concluding Statement*, *in* SOCIAL MEDIA, FREEDOM OF SPEECH AND THE FUTURE OF OUR DEMOCRACY 327, 328 (Lee C. Bollinger & Geoffrey R. Stone eds., 2022) (advocating for a system of government oversight that forbids censorship but enables government regulation to protect the public interest).

87 Tim Wu identifies different assumptions about the First Amendment and argues that informational scarcity, the listeners' abundant time and interest, and the

technologies and the resulting new means of manipulation, the digital marketplace of ideas is broken.[88]

Filter bubbles place an almost insurmountable burden on counterspeech: truth does not win in an exchange of ideas and opinions, but misinformation remains unchallenged.[89] The idea that more speech is the appropriate remedy for bad speech does not work if that speech does not reach the original speaker or only reinforces their views.[90] The U.S. government's efforts to engage in counterspeech through the Global Engagement Center have proven to be mostly ineffective.[91] The failure of counterspeech in the digital realm can also be attributed to the highly emotional nature of social media. Constitutional democracy can only speak rationally, which is unlikely to address emotional claims.[92] "[E]motionalism is inherent in democracy, but democracy, especially in the form of representative government, was designed as a characteristically non-emotional

government's use of criminal law or other coercive means are the main threats to the First Amendment. Wu, *supra* note 21, at 553–54. However, he also argues that all three of these things became obsolete due to the latest changes in communication technologies. *Id.* at 554.

88 Some argue that the First Amendment is way too "Lochnerian" in the first place, meaning that it misbalances all involved interests too much in favor of the speaker. *See, e.g.*, JOHN RAWLS, POLITICAL LIBERALISM 362–63 (1993). *But see* Genevieve Lakier, *The First Amendment's Real Lochner Problem*, 87 U. CHI. L. REV. 1241, 1241 (2020) (criticizing the focus of the widespread academic criticism on its economic aspects); *see also* William French, *This Isn't Lochner, It's the First Amendment: Reorienting the Right to Contract and Commercial Speech*, 114 NW. U.L. REV. 469, 472–74, 490–501 (2019) (addressing the blurred boundaries of the First Amendment and Lochnerianism).

89 *See* Anna Rhoads, *Incitement and Social Media-Algorithmic Speech: Redefining* Brandenburg *for a Different Kind of Speech*, 64 WM. & MARY L. REV. 525, 544 (2022); *see also* Napoli, *supra* note 34, at 77–79 (addressing the failure of social media to provide counterspeech due to echo chambers).

90 *Cf.* SUNSTEIN, *supra* note 30, at 88.

91 *See, e.g.*, Neill Perry, *The Global Engagement Center's Response to the Coronavirus Infodemic*, CYBER DEF. REV., Spring 2022, at 131, 132 (stating that the Global Engagement Center is "ill-suited for disinformation intended for American audiences"); OFF. OF INSPECTOR GEN., INSPECTION OF THE GLOBAL ENGAGEMENT CENTER 7–9 (2022) (finding that the Global Engagement Center's effectiveness was limited).

92 *See* Karl Loewenstein, *Militant Democracy and Fundamental Rights, I*, 31 AM. POL. SCI. REV. 417, 428 (1937); *see also* Sajó, *supra* note 5, at 202; András Sajó, *Militant Democracy and Emotional Politics*, 19 CONSTELLATIONS 562, 569 (2012). An attempt to meet populist emotionalism is sometimes seen in the concept of *Verfassungspatriotismus*, or constitutional patriotism. For more on this topic, see *SVETLANA* TYULKINA, MILITANT DEMOCRACY: UNDEMOCRATIC POLITICAL PARTIES AND BEYOND (2015).

institution."[93] Therefore, the concept of counterspeech can only be applied in a limited way to inherently emotional social media.[94]

As a result of misinformation, the former marketplace is now being (often intentionally) flooded with false information for the sole purpose of confusing public opinion. The question that is being rightly asked here is to what extent misinformation is contributing to the marketplace of ideas.[95] It becomes virtually impossible to distinguish between true and false information.[96] This makes it easy for people with more resources to manipulate the online marketplace of ideas.[97] The need to counter misinformation is also reflected in the international arena.[98]

Due to bots, certain speech seems far more persuasive because of the perceived mass of its followers.[99] Hate speech can lead to the targeted silencing of vulnerable individuals and have a chilling effect on the greater public.[100] In response, social media platforms are imposing restrictions to tackle the issue of hate speech.[101] This raises many First Amendment issues.[102]

The clear and present danger test is not fit to address these threats.[103] The physical and temporal remoteness that the

---

93 Sajó, *supra* note 92, at 563.

94 *See* David A. Strauss, *Social Media and First Amendment Fault Lines*, *in* SOCIAL MEDIA, FREEDOM OF SPEECH AND THE FUTURE OF OUR DEMOCRACY 3, 15 (Lee C. Bollinger & Geoffrey R. Stone eds., 2022).

95 Therefore, the type of disinformation that Henricksen calls "fraudulent" speech should not be protected by the First Amendment. *See* Henricksen, *supra* note 40, at 556, 558.

96 *See* Napoli, *supra* note 34.

97 *See* Netanel, *supra* note 22, at 507–10. This, of course, is no unique feature of the social media sphere. However, the characterizing technologies at play allow for a far more powerful manipulation of the marketplace of ideas than would be possible in the analog world.

98 *See, e.g.*, *Reykjavík Declaration*, COUNCIL OF EUR. 16 (May 17, 2023), https://coebank.org/documents/1373/4th_CoE_Summit_Reykjavik_Declaration.pdf [https://perma.cc/Q32V-5S8J] (highlighting the belief of the Council of Europe that disinformation and misinformation pose a threat to democracy that must be countered).

99 *See* SUNSTEIN, *supra* note 30, at 74–75 (discussing the phenomenon of group polarization and "the effects of social media platforms on which people's views end up being constantly reaffirmed by like-minded types").

100 *See* Netanel, *supra* note 22, at 506–10; Wu, *supra* note 21, at 564–65 (noting examples of such); *see also* FISS, *supra* note 70, at 15–18 (proposing that we ought not to think about freedom of speech without considering the equality of the different speakers).

101 *See* Nunziato, *supra* note 62, at 1538–54 (detailing what Facebook and X (formerly Twitter) are doing to regulate speech).

102 Dawn C. Nunziato, *The Death of the Public Forum in Cyberspace*, 20 BERKELEY TECH. L.J. 1115, 1170–71 (2005) (criticizing the privatization of the digital world and the accompanying erosion of First Amendment values).

103 *See* Amélie Heldt, *Terror-Propaganda Online: Die Schranken der Meinungsfreiheit in Deutschland und den USA*, 2017 NEUE JURISTISCHE ONLINE-ZEITSCHRIFT 1458, 1460–61 (comparing the *Brandenburg* doctrine with the German doctrine).

internet creates, as the feed sometimes displays posts long after they are posted, makes it difficult to meet the imminence requirement.[104] Moreover, many of the aforementioned threats to democracy are simply too abstract to be addressed by the clear and present danger test.[105] Throwing one's hands in the air and letting the platforms self-police cannot be the solution.[106]

Considering all these factors, it becomes apparent that there is no marketplace of ideas on social media, at least none of the kind envisioned by the First Amendment.[107] Thus, the vaunted marketplace is experiencing a "market failure" on social media.[108]

III. MILITANT DEMOCRACY FOR THE FIRST AMENDMENT?

Many recognize that a shift in First Amendment doctrine is needed.[109] The concept of militant democracy provides an

---

104 *See* Rhoads, *supra* note 89, at 543. On the other hand, some argue that the clear and present danger test protects too little speech. *See* David R. Dow, *The Moral Failure of the Clear and Present Danger Test*, 6 WM. & MARY BILL RTS. J. 733, 733 (1998) (asserting that "evil words do not always lead to evil acts"). Interestingly, Meiklejohn has already doubted the "present" requirement of the clear and present danger test because he considers it to be arbitrary and exploitable. *See* MEIKLEJOHN, *supra* note 60, at 47–49.

105 Meiklejohn strongly criticized the clear and present danger test as unfitting for American democracy, saying that the effect of this doctrine "upon our understanding of self-government has been one of disaster." MEIKLEJOHN, *supra* note 60, at 33.

106 Skeptical about the adequacy of *Brandenburg* as it pertains to social media, Connie Hassett-Walker appears to trust the self-policing of online platforms. Connie Hassett-Walker, *Does Brandenburg v. Ohio Still Hold in the Social Media Era? Racist (and Other) Online Hate Speech and the First Amendment*, 8 COGENT SOC. SCIS., Feb. 18, 2022, at 1, 14–15.

107 Interestingly, Meiklejohn was already "bitterly disappointed" by how the radio—a new form of communication at the time—was not a free space for such an exchange of ideas and that it was instead more engaged in making money. MEIKLEJOHN, *supra* note 60, at 87. Meiklejohn saw this as a danger for democracy, noting that this new form of media was "not cultivating those qualities of taste, of reasoned judgment, of integrity, of loyalty, of mutual understanding upon which the enterprise of self-government depends." *Id.* However, he also noted, "On the contrary, it is a mighty force for breaking them down." *Id.*

108 Napoli, *supra* note 34, at 88–93.

109 *See, e.g.*, Bollinger & Stone, *supra* note 86, at 328–29; Larry Kramer, *A Deliberate Leap in the Opposite Direction: The Need to Rethink Free Speech*, *in* SOCIAL MEDIA, FREEDOM OF SPEECH AND THE FUTURE OF OUR DEMOCRACY 17, 20 (Lee C. Bollinger & Geoffrey R. Stone eds., 2022); Mark S. Kende, *Social Media, the First Amendment, and Democratic Dysfunction in the Trump Era*, 68 DRAKE L. REV. 273, 274–75 (2020); Wu, *supra* note 21, at 581; Jill I. Goldenziel & Manal Cheema, *The New Fighting Words?: How U.S. Law Hampers the Fight Against Information Warfare*, 22 U. PA. J. CONST. L. 81, 167–70 (2019) (arguing that the current doctrine poses a national security risk); Kenneth Propp, *Speech Moderation and Militant Democracy: Should the United States Regulate like Europe Does?*, ATL. COUNCIL (Feb. 1, 2021), https://www.atlanticcouncil.org/blogs/new-atlanticist/speech-moderation-and-militant-democracy-should-the-united-states-regulate-like-europe-does/ [https://perma.cc/B8T3-FBJT] (calling "adjustments at the margins . . . desirable" from a foreign affairs perspective).

alternative line of argument for defending democratic institutions against the harms of social media,[110] and it has a chance of addressing those new challenges and challengers.[111]

The following section will introduce militant democracy into the discussion as a solution for a change in First Amendment doctrine.[112]

A. The Concept

1. Origins

"[M]ilitant democracy can be defined as the capacity of liberal democracies to defend themselves against challenges to their continued existence by taking pre-emptive action against those who want to overturn or destroy democracy by abusing democratic institutions and procedures."[113] The basic ideas for the political concept of militant democracy were developed by Karl Loewenstein and Karl Mannheim, scholars who both fled Germany during the Nazi era.[114]

In 1937, Loewenstein developed the model of militant democracy against the background of his experiences with National Socialism (the ideology of Hitler's Nazi regime).[115] The trigger for Loewenstein's work was the new use of emotionalism in politics. He feared that "[t]he technical devices for mobilizing emotionalism" would lead to a more emotional public debate.[116] He saw firsthand how fascism and populism "exploit the tolerant confidence of democratic ideology that in the long run truth is stronger than falsehood."[117] Because of this experience, he was

---

110 Netanel, *supra* note 22, at 494.

111 Angela K. Bourne & Bastiaan Rijpkema, *Militant Democracy, Populism, Illiberalism: New Challengers and New Challenges*, 18 EUR. CONST. L. REV. 375, 378–79 (2022) (examining both the structural change that Militant Democracy itself must endure due to new challenges, and whether Militant Democracy is a good fit to combat populist actors with huge support).

112 For an introduction of Militant Democracy already meeting social media harms, see Aziz Z. Huq, *Militant Democracy Comes to the Metaverse?*, 72 EMORY L.J. 1105, 1124–26, 1135–37 (2023); Netanel, *supra* note 22, at 557–59; *see also* Thomas M. Keck, *Erosion, Backsliding, or Abuse: Three Metaphors for Democratic Decline*, 48 L. & SOC. INQUIRY 314, 332 (2023).

113 TYULKINA, *supra* note 92, at 15.

114 *See* Loewenstein, *supra* note 92, at 421, 423–24, 428–29; KARL MANNHEIM, DIAGNOSIS OF OUR TIME: WARTIME ESSAYS OF A SOCIOLOGIST 4–8 (1943). For a discussion on how the idea of defending democracy against its potential enemies traces back much further than the 1930s, see TYULKINA, *supra* note 92, at 14.

115 Lowenstein, *supra* note 92, at 417.

116 *Id.* at 418.

117 *Id.* at 424.

convinced that legislative countermeasures should meet the emotional technique.[118]

Loewenstein is often associated with the call for party bans.[119] Although party bans are part of militant democracy in many countries,[120] they are highly controversial due to the danger of misuse and the position of parties at the heart of a representative democracy.[121] But Loewenstein's understanding of militant democracy goes much further. It predominantly focuses on preventing fascists from undermining democracy by disrupting democratic discourse.[122] This can be achieved through party bans, but also by a variety of other means.

Karl Mannheim's thoughts on militant democracy were influenced by his experiences with Nazi Germany as well. He saw militant democracy as a third option between laissez-faire liberalism and totalitarian dictatorship,[123] which is "said to be the manifestation of his standpoint as a defender of freedom and democracy against fascism and totalitarianism."[124] Although their understanding of democracy is sometimes described as elitist[125] and undemocratic,[126] these two thinkers were

---

118 *Id.* at 431.

119 *See* Karl Loewenstein, *Militant Democracy and Fundamental Rights, II*, 31 AM. POL. SCI. REV. 638, 646–48 (1937). For an extensive analysis of and critical remarks on the party ban practice in Europe, see generally Angela K. Bourne & Fernando Casal Bértoa, *Mapping 'Militant Democracy': Variation in Party Ban Practices in European Democracies (1945-2015)*, 13 EUR. CONST. L. REV. 221, 221–25 (2017).

120 *See, e.g.*, *Proceedings for the Prohibition of a Political Party*, BUNDESVERFASSUNGSGERICHT FED. CONST. CT., https://www.bundesverfassungsgericht.de/EN/TheFederalConstitutionalCourt/TypesOfProceedings/ProceedingsForTheProhibitionOfAPoliticalParty/proceedingsfortheprohibitionofapoliticalparty_node.html [https://perma.cc/W5TY-AKJW] (last visited Apr. 9, 2025) (explaining the proceedings for party bans in Germany's Federal Constitutional Court).

121 *See* Peter Stone, *Democratic Equality and Militant Democracy*, *in* MILITANT DEMOCRACY AND ITS CRITICS 38, 45 (Anthoula Malkopoulou & Alexander S. Kirshner eds., 2019).

122 *See* Loewenstein, *supra* note 119, at 651–53. For an overview of creative, "softer" militant democracy measures, see TYULKINA, *supra* note 92, at 110–19.

123 MANNHEIM, *supra* note 114, at 7.

124 Ryusaku Yamada, *Mannheim, Mass Society and Democratic Theory*, *in* THE ANTHEM COMPANION TO KARL MANNHEIM 51, 58 (David Kettler & Volker Meja eds., 2018).

125 *See id.* at 52. On the legitimacy problem of Militant Democracy, see Cristóbal Rovira Kaltwasser, *Militant Democracy Versus Populism*, *in* MILITANT DEMOCRACY AND ITS CRITICS 72, 85–86 (Anthoula Malkopoulou & Alexander S. Kirshner eds., 2019).

126 This is particularly due to Mannheim's emphasis on the role of science and expertise, as well as his distrust of the electorate. *See* Martyn Hammersley, *Karl Mannheim on Fascism: Sociological Lessons About Populism and Democracy Today?*, 28 SOCIO. RSCH. ONLINE 320, 327 (2023) (citing FRIEDRICH A. HAYEK, THE ROAD TO SERFDOM (1944)). Mannheim has been described as "neither liberal nor democratic," and is accused of not trusting the people as the true sovereign in a democracy. *Id.*

undoubtedly right to recognize that the greatest dangers to democracy come from within. The recognition that democracy must be protected from destructive forces from within is important and rightly still influential today.[127]

The fact that this approach contains formally anti-democratic elements does not necessarily constitute a problem within a constitutional democracy. Constitutional safeguards often appear anti-democratic and counter-majoritarian in the short term, while realistically they are designed to protect constitutional democracy in the long run.[128] Because of their effect, populists often characterize constitutional safeguards as elitist and illegitimate obstacles to popular power.

2. Militant Democracy in Germany and Europe

To understand what militant democracy can look like, it is helpful to examine its specific forms and implementations in the European and German legal systems.[129] A variety of examples will be provided below, but the focus will be on examples related to freedom of opinion.

Germany is a paradigmatic case of militant democracy, largely because this principle is in many ways reflected in the German constitution, the Basic Law. The obvious background for this strong emphasis on the importance of a self-protecting

---

127 Later, many other thinkers introduced their understandings of Militant Democracy. One of the most influential was Max Lerner. On the differences between Loewenstein's and Lerner's doctrines, see Graham Maddox, *Karl Loewenstein, Max Lerner, and Militant Democracy: An Appeal to 'Strong Democracy,'* 54 AUSTL. J. POL. SCI. 490, 491–96 (2019).

128 An obvious example is the two-term limit for the presidency in the Twenty-Second Amendment to the Constitution. U.S. CONST. amend. XXII, § 1. This limitation, sparked by worries of executive tyranny, is clearly intended to secure democracy in the long term and to prevent despotism. *See, e.g.*, Mark Satta, *Why Does the U.S. Have Presidential Term Limits? The History of the 22nd Amendment*, PBS NEWS (Apr. 5, 2025, 2:30 PM), https://www.pbs.org/newshour/politics/why-does-the-u-s-have-presidential-term-limits-the-history-of-the-22nd-amendment [https://perma.cc/HF3B-AR2X]. Nevertheless, it is formally undemocratic, denying voters and potential candidates the chance to align with the popular will by electing a president for a third term. The same is true for many forms of minority protection. The idea that the minority has the ability to become the majority in the long run is deeply important for democracy. Yet, enacting protections to ensure there is a possibility of this result is formally antidemocratic in the first place. This dichotomy is inherent in every constitutional democracy.

129 In the United States, Militant Democracy is often perceived as a European or German concept. It is frequently accompanied by the opinion that there is limited freedom of speech in Europe and especially in Germany. *See, e.g.*, Ronald J. Krotoszynski, Jr., *A Comparative Perspective on the First Amendment: Free Speech, Militant Democracy, and the Primacy of Dignity as a Preferred Constitutional Value in Germany*, 78 TUL. L. REV. 1549, 1597 (2004).

democracy is, of course, the failure of the Weimar Republic. Germany has experienced firsthand how easily a democracy can abolish itself. The Nazis made no secret of their plan to defeat democracy with their own weapons. Joseph Goebbels, the Nazi Party's chief propagandist and Reich Minister of Propaganda, famously wrote:

> We enter the Reichstag to arm ourselves with democracy's weapons. If democracy is foolish enough to give us free railway passes and salaries, that is its problem. It does not concern us. Any way of bringing about the revolution is fine by us. . . . We are coming neither as friends or neutrals. We come as enemies! As the wolf attacks the sheep, so come we.[130]

Based on this experience, the framers of the Basic Law sought to prevent history from repeating itself.[131] Famously, the German Federal Constitutional Court stated that the "Basic Law can be largely interpreted as an antithesis to the totalitarianism of the National Socialist regime, and . . . seeks to learn from historical experience and to rule out a repeat of such injustice once and for all."[132] Prominent examples of this militant constitutionalism are the party ban procedure[133] and the withdrawal of basic (political) rights.[134] These measures are probably the most severe manifestations of militant democracy in the Basic Law. However, they have been used only sparingly or

---

130 JOSEPH GOEBBELS, DER ANGRIFF: AUFSÄTZE AUS DER KAMPFZEIT 71–73 (1936), *translated in* Randall Bytwerk, *Why Do We Want to Join the Reichstag?*, CALVIN U., https://research.calvin.edu/german-propaganda-archive/angrif06.htm [https://perma.cc/4VCK-452N].

131 For an extensive discussion on the various interpretations of the "Never Again" doctrine in German constitutional law, see Daniel Wolff, *Zeit und Verfassung – Konstitutionelle Reaktionen auf den Zeitregimewechsel* (Deutsche Forschungsgemeinschaft, Project Identification No. 558486849, 2024).

132 BVerfG, 1 BvR 2150/08, Nov. 4, 2009, para. 42, https://www.bundesverfassungsgericht.de/SharedDocs/Entscheidungen/EN/2009/11/rs20091104_1bvr215008en.html [https://perma.cc/Z7PM-N4L8] (Ger.).

133 "Parties that, by reason of their aims or the behaviour of their adherents, seek to undermine or abolish the free democratic basic order or to endanger the existence of the Federal Republic of Germany shall be unconstitutional." Grundgesetz [GG] [Basic Law], art. 21, para. 2, translation at https://www.gesetze-im-internet.de/englisch_gg/englisch_gg.html [https://perma.cc/V6Z5-W7F7] (Ger.). The German Constitutional Court must determine the unconstitutionality in each individual case. *Id.* para. 4.

134 Grundgesetz [GG] [Basic Law], art. 18, translation at https://www.gesetze-im-internet.de/englisch_gg/englisch_gg.html [https://perma.cc/4CEE-V6RH] (Ger.). Article 18 provides that anyone who abuses certain political rights to fight against the free democratic basic order forfeits those rights. *Id.*

not at all.[135] More relevant in practice is the general hierarchy in the Basic Law, which is established above all by the Perpetuity Clause.[136] This clause prevents the most important principles from being changed through constitutional amendment, including human dignity,[137] democracy, and the rule of law, among others.[138] This hierarchy plays a major role in legal interpretation.

In Germany, as in the United States, the role of free speech is seen as twofold, balancing individual freedom with its role in democracy.[139] However, the link between democracy and free speech is more established in German constitutional doctrine than in the United States. In German constitutional law, fundamental rights do not merely have a defensive, liberal character, but can also represent an objective set of values or establish duties for the state to protect the constitutionally guaranteed freedoms.[140] This is underscored by a strong belief that an informed public is important and even essential for democracy.[141] Related to this understanding is the conviction that communication spaces must be inclusive and deliberative, even in the digital space, because only widespread participation in the political debate—which is based on undisputed facts as a common ground—can counteract the fragmentation and polarization of society.[142] True to this philosophy, speech that

---

135 So far there have been two party bans—one in 1952 against the Socialist Reich Party, which drew on the legacy of the National Socialist regime, and another in 1956 against the Communist Party of Germany—but there has been no withdrawal of basic rights. *See Banning Political Parties*, FED. MINISTRY OF THE INTERIOR, https://www.bmi.bund.de/EN/topics/constitution/law-political-parties/banning-pol-parties/banning-pol-parties-node.html [https://perma.cc/Y7ZM-T9QG] (last visited May 8, 2025).

136 Grundgesetz [GG] [Basic Law], art. 79, para. 3, translation at https://www.gesetze-im-internet.de/englisch_gg/englisch_gg.html [https://perma.cc/4CEE-V6RH] (Ger.).

137 *Id.* art. I.

138 *Id.* art. 20, para. 1.

139 *See* VIKTOR VOLKMANN, MEINUNGSFREIHEIT FÜR DIE FEINDE DER FREIHEIT? 275 (2019).

140 For one of the most famous cases of the German Federal Constitutional Court on this issue, see Bundesverfassungsgericht [BVerfG] [Federal Constitutional Court] Jan. 15, 1958, 7 Entscheidungen des Bundesverfassungsgerichts [BVerfGE] 198 (Ger.).

141 For example, the German Federal Constitutional Court calls *Rundfunkgebühren* (the public broadcasting fee) a *Demokratieabgabe* (levy on democracy). BVerfG, 1 BvR 1675/16, July 18, 2018, https://www.bundesverfassungsgericht.de/SharedDocs/Entscheidungen/EN/2018/07/rs20180718_1bvr167516en [https://perma.cc/QVN5-424G]. On the applicability of this decision to social media, see Michael Fehling & Matthias Leymann, *Der Neue Strukturwandel der Öffentlichkeit: Wie Lassen Sich die Sozialen Medien Regulieren?*, 51 AFP 110, 111 (2020).

142 *Cf.* Cristina Lafont, *Deliberative Demokratie Nach der Digitalen Transformation*, BPB (Oct. 20, 2023), https://www.bpb.de/shop/zeitschriften/apuz/diskurskultur-

contributes to public democratic discourse is most strongly protected. The conscious statement of false facts is not protected because it does not contribute to the formation of an opinion on an accurate factual basis. There is a strong belief that democracy must guarantee the conditions for democratic participation.[143] The goal is not to eradicate bad ideas, but simply to ensure that they are not presented in a historically false or misleading way.[144]

Because of the rise of the political far right in Germany, some might criticize the German system of militant democracy as a failure.[145] To a certain extent, this argument is fair. However, it fails to recognize that the global trend of democratic backsliding is, at least for now, less dramatic in Germany than in many other democracies.

A similar pattern can be observed recently at the level of the European Union (EU).[146] The EU is highly dependent on its states being democratic.[147] With the European Commission's 2020 European Democracy Action Plan, the EU is demonstrating

---

2023/541846/deliberative-demokratie-nach-der-digitalen-transformation/ [https://perma.cc/EPJ3-V7UR].

143 On the inherent difficulty of the liberal, secularized state to guarantee its own prerequisites, see Ernst-Wolfgang Böckenförde, *Die Entstehung des Staates als Vorgang der Säkularisation*, *in* SÄKULARISATION UND UTOPIE 75, 75–78 (Sergius Buve ed., 1967). The former judge of the German Constitutional Court, Ernst-Wolfgang Böckenförde, established that "[t]he liberal, secularized state is sustained by conditions it cannot itself guarantee." Mirjam Künkler & Tine Stein, *Böckenförde on the Secular State and Secular Law*, *in* 2 RELIGION, LAW, AND DEMOCRACY 138, 139 (Mirjam Künkler & Tine Stein eds., 2020) (citation omitted). Therefore, the liberal state which abolished absolutism and established the sovereignty of the people could not impose norms like the notion of loyalty to liberty and democracy on the population without in itself being absolutist again. *See id.* at 139–45.

144 For a discussion on the issue of falsifying history in the United States, see Henricksen, *supra* note 40, at 570–72. On Germany's fight against misinformation and hate speech on social media with a special focus on the Network Enforcement Act (in German, *Netzwerkdurchsetzungsgesetz*, or NetzDG for short), see Nunziato, *supra* note 62, at 1532–38; *see also* Finnegan, *supra* note 51, at 612 (explaining that the NetzDG would not pass First Amendment scrutiny).

145 *See* Krotoszynski, *supra* note 129, at 1598–99.

146 Ruti Teitel argues that today's acceptance "of a robust conception of the rule of law" has allowed for a more nuanced approach "to the balance of values" in the European arena. Ruti Teitel, *Militating Democracy: Comparative Constitutional Perspectives*, 29 MICH. J. INT'L. L. 49, 70 (2007).

147 *Cf.* Tore Vincents Olsen, *Liberal Democratic Sanctions in the EU*, *in* MILITANT DEMOCRACY AND ITS CRITICS 150, 151–55 (Anthoula Malkopoulou & Alexander S. Kirshner eds., 2019). In addition, not only the EU, but many international organizations are highly dependent on a certain level of democracy and law-abiding behavior of their member states. *See* Constanze Stelzenmüller, *Democracy: Transatlantic Action Plan*, DGAP (Jan. 19, 2021), https://dgap.org/en/research/publications/democracy [https://perma.cc/DR54-QW4X] (discussing NATO's reliance on member states' commitment to democracy).

militant democracy at the intergovernmental level.[148] At the same time, the Digital Services Act (DSA) provides an example on what militant democracy could look like online, in this instance by imposing regulations on social media platforms.[149] The DSA would arguably fail First Amendment scrutiny.[150] However, some hope that the DSA will have a serious impact on the social media landscape in the United States due to the "Brussels Effect"—the phenomenon whereby EU regulations influence global business practices because international companies adopt them to maintain access to the European market—and that even if the neoliberal U.S. policy approach does not change, it could push social media in the United States to become a more reasoned and fact-based public sphere.[151] It is also worth noting that there seems to be a hesitant tendency to reflect the concept of militant democracy in public international law.[152] Militant democracy is not only shaping national but also supranational and transnational government frameworks in response to the evolving challenges of the digital age.

---

148 *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: On the European Democracy Action Plan*, at 1, COM (2020) 790 final (Dec. 3, 2020). For an analysis of the related enforcement mechanisms, see Olsen, *supra* note 147, at 157–66; *see also* Netanel, *supra* note 22, at 496.

149 *See* Netanel, *supra* note 22, at 559–61.

150 *Id.* at 575.

151 *See id.* at 496–97. That said, this hope seems to show an abandonment of any chance to resolve this issue through domestic legislation. *See id.* at 493, 496.

152 *See* TYULKINA, *supra* note 92, at 51–54; *see also* Netanel, *supra* note 22, at 552–55. The Council of Europe itself could be seen as the embodiment of Militant Democracy on an international level, trying to secure an international democratic community to protect human rights. *Cf.* Convention for the Protection of Human Rights and Fundamental Freedoms, pmbl., Nov. 4, 1950, 213 U.N.T.S. 221.

> Reaffirming their profound belief in those fundamental freedoms which are the foundation of justice and peace in the world and are best maintained on the one hand by an effective political democracy and on the other by a common understanding and observance of the Human Rights upon which they depend. . . .

Id. However, conflicts frequently occur between the European Convention on Human Rights (ECHR) and domestic Militant Democracy efforts. *See* TYULKINA, *supra* note 92, at 95–102. Still, the ECHR seems to recognize Militant Democracy's international legality. *See* Patrick Macklem, *Militant Democracy, Legal Pluralism, and the Paradox of Self-Determination*, 4 INT'L J. CONST. L. 488, 516 (2006). For insight into the internationalization of Militant Democracy in connection with international human rights law, see generally Christian Walter, *Interactions Between International and National Norms: Towards an Internationalized Concept of Militant Democracy*, *in* MILITANT DEMOCRACY—POLITICAL SCIENCE, LAW AND PHILOSOPHY 79, 79–94 (Afshin Ellian & Bastiaan Rijpkema eds., 2018). Finally, on the problems that come with international or supranational forms of democracy protection, see Jan-Werner Müller, *Protecting Popular Self-Government from the People? New Normative Perspectives on Militant Democracy*, 19 ANN. REV. POL. SCI. 249, 263 (2016).

B. Applicability to the United States

Having established what militant democracy can look like, it is now time to turn to the question of whether the U.S. Constitution is compatible with a militant democracy-type approach. But what is the appropriate judging metric? Given that U.S. constitutional law is currently dominated by originalism and textualism, it seems sensible to look to the text and history of the U.S. Constitution, as well as to general constitutional philosophy and doctrine, for sources of constitutional self-defense.

1. Militant Democracy and the U.S. Constitution

In trying to answer how a constitution can be militant,[153] one should consider whether it is concerned with its survival. This Article will try to identify such a concern by looking at whether the text and history of the U.S. Constitution indicate that the Founders were indeed afraid that this form of democratic government might pose a threat to itself and whether they put in place certain safeguards.

Without even looking at the U.S. Constitution, it could be argued that any constitutional democracy that seeks to preserve itself and prevent despotism is, at least on an abstract level, already inherently *militant*.[154] Many of the democratic institutions could be seen as preventive per se and are, on an abstract level, again protected by constitutional limitations on the government's ability to change these institutions.[155]

Such indicators can be found in the U.S. Constitution as well, from the system of checks and balances to the two-term limit on the presidency to the protections of the Bill of Rights. More concrete examples are: (1) the high procedural requirements for constitutional amendments;[156] (2) term limits;[157] (3) a fundamental skepticism of direct democracy,[158] which is more susceptible to populism,[159] and therefore limited in

153 On the issue of militant constitutions, see generally Sajó, *supra* note 5, at 187–203.

154 *See* Sajó, *supra* note 5, at 188; *see also* Svetlana Tyulkina, *Militant Democracy as an Inherent Democratic Quality*, *in* MILITANT DEMOCRACY AND ITS CRITICS 207, 207 (Anthoula Malkopoulou & Alexander S. Kirshner eds., 2019).

155 *See* Sajó, supra note 5, at 188.

156 On this requirement as well as on its weak points, see *id.* at 195–96.

157 *Id.* at 196.

158 *See, e.g.*, THE FEDERALIST NO. 10 (James Madison).

159 On the danger of populism, see *id.*

its role;[160] (4) the multi-layered sovereignty represented by federalism; and (5) the delegation of power to international organizations.[161] Ultimately, the whole concept of protecting minorities to prevent a tyranny of the majority is inherently preventive.[162] One could even argue that the very purpose of any constitutional institution and of any constitution itself is not only to establish the government but also to limit popular power.[163] The U.S. Supreme Court recognized as early as *Marbury v. Madison* that the judiciary must check executive and legislative power out of sheer logical necessity.[164] Therefore, "[p]reventive constitutionalism *can* be justified within liberal constitutional theory as self-preservation."[165]

However, there are two forms of militant constitutionalism. While the first seeks to limit the possibility of enemies of democracy coming into power, the second intends to build institutional resilience for the time after they have taken power.[166] Most of the aforementioned examples fall under the latter form of Institutional militant constitutionalism,[167] as they aim to build institutional resilience, especially in the rule of law.[168] But to determine whether militant democracy can be used as an argument for a change in First Amendment doctrine, the question is not one of institutional defensiveness, but one of preventiveness.[169] One must therefore move away from abstract

---

160 *See* Sajó, *supra* note 5, at 197.

161 This is sometimes referred to as a vertical separation of powers. For a discussion on the EU as a barrier to illiberalism, see *id.* at 198.

162 On the question of whether a populist, "shallow," and illiberal democracy is to be considered an illegitimate form of government, see *id.* at 189–93.

163 As the prime example of this concept, see THE FEDERALIST NO. 51 (James Madison). *See also* Sajó, *supra* note 5, at 194–95.

164 *See* Marbury v. Madison, 5 U.S. (1 Cranch) 137, 177–78 (1803).

165 Sajó, *supra* note 5, at 203.

166 *See id.* at 193–94. Another way to categorize different models of Militant Democracy is to distinguish between militant, procedural, and social self-defense. Anthoula Malkopoulou & Ludvig Norman, *Three Models of Democratic Self-Defence*, *in* MILITANT DEMOCRACY AND ITS CRITICS 92, 92–93 (Anthoula Malkopoulou & Alexander S. Kirshner eds., 2019).

167 For examples on different "institutional instruments" of Militant Democracy, see Tom van der Meer & Bastiaan Rijpkema, *Militant Democracy and the Minority to Majority Effect: On the Importance of Electoral System Design*, 18 EUR. CONST. L. REV. 511, 512–513 (2022) (analyzing the role of proportional representation of the population in the electorate).

168 A new frame for a part of this institutional militant democracy can be seen in the call for a "militant rule of law." András Sajó, *Militant Rule of Law*, VERFASSUNGSBLOG (Dec. 20, 2023), https://verfassungsblog.de/militant-rule-of-law/ [https://perma.cc/YKK5-QQG6].

169 The term "preventive constitutionalism" is also used by András Sajó. Sajó, *supra* note 5, at 193, 198.

considerations and take a closer look at the text and history of the U.S. Constitution to examine whether the Founders recognized the dangers to democracy from within, and whether they saw the need to establish certain preventative safeguards.

a. Reconstruction Era

The Reconstruction era from 1863 to 1877 is perhaps the most obvious example of preventiveness in the history of the U.S. Constitution. In fact, Professor Alexander S. Kirshner calls Reconstruction a "Paradigmatic Case of Militant Democracy."[170] Reconstruction aimed to protect the Union from being undermined by former Confederates who still posed a major threat to its existence. Fearful of losing to the Union's declared enemies in Congress after finally winning the war, former Confederate members of the House and Senate were politically excluded.[171] But more importantly, the Fourteenth Amendment of the U.S. Constitution, with its Insurrection Clause in Section 3, was adopted.[172] The Insurrection Clause is designed to block enemies of the United States from coming to power. The entire Reconstruction era, and especially its enduring manifestation in the Fourteenth Amendment, was clearly intended to protect democracy from actors who would use democracy to attack and destroy it. Kirshner understands Reconstruction as "an indispensable model for the self-limiting defense of democracy."[173] Intending to secure democracy, the Reconstructionist approach "was realistically aimed at securing a polity that was both reasonably democratic and legitimate."[174] Kirshner compares non-democratic players to a chess player who does not abide by the rules and thus excludes himself.[175] This concept of "democratically conditional exclusion"[176] thus appears to be represented in the Constitution. Whoever wants to play a

170 ALEXANDER S. KIRSHNER, A THEORY OF MILITANT DEMOCRACY: THE ETHICS OF COMBATTING POLITICAL EXTREMISM 144 (2014).

171 *Id.* at 142.

172 *See* Keck, *supra* note 112, at 334 (arguing that the Insurrection Clause is an early manifestation of Militant Democracy-like ideas); *see also* Mark A. Graber, *Who's Afraid of Militant Democracy, U.S. Style*, VERFASSUNGSBLOG (Feb. 20, 2024), https://verfassungsblog.de/whos-afraid-of-militant-democracy-u-s-style/ [https://perma.cc/MMW2-RCJ6].

173 KIRSHNER, *supra* note 170, at 143.

174 *Id.*

175 *Id.* at 155–56.

176 *Id.* at 152.

role in the political game must not be an enemy of the game itself and must not pose a major threat to the game.[177]

While considering the Insurrection Clause, one must note that its actual relevance seems to be limited. First, extremists no longer conspire in large numbers for open revolution or insurrection. After their failed overthrow attempts in the twentieth century, they have learned that it is easier to destroy a democracy from within.[178] Second, the recent decision of *Trump v. Anderson*, which requires legislation to enforce Section 3, seems to tragically diminish the practical role of the Insurrection Clause.[179] Unfortunately, the majority in *Trump* does not address the democracy-preserving function of the insurrection clause at all. Justices Sonia Sotomayor, Elena Kagan, and Ketanji Brown Jackson, however, recognize its "important . . . role in our democracy"[180] and criticize the majority for disregarding this purpose of the Insurrection Clause, arguing that it would "insulate all alleged insurrectionists from future challenges to their holding federal office."[181]

b. Not a Suicide Pact

The belief that more speech leads to more freedom and democracy dominates the First Amendment doctrine. True to this notion, the number of opinions expressed correlates with the functioning of democratic discourse; the more views expressed, the better they interact in the marketplace of ideas and promote democracy. This strong sense of inseparability of democracy and freedom of speech has led to First Amendment absolutism.

Such absolutism, however, has risen to extreme formalism and has forgotten its roots. Alexander Meiklejohn already warned that an overly formalistic approach to new forms of

---

177 John Rawls also supported the idea that the "liberal state will not concede the space of politics to those who want to use that space to destroy it." David Dyzenhaus, *Legal Theory in the Collapse of Weimar: Contemporary Lessons?*, 91 AM. POLIT. SCI. REV. 121, 121 (1997). Rawls was convinced that if the Constitution is not secure, there may be a reason to deny certain freedoms to the intolerant. *See* JOHN RAWLS, A THEORY OF JUSTICE 337 (rev. ed. 1999). In this context, some scholars call to abridge the absoluteness which dominates First Amendment doctrine. *See* SUNSTEIN, *supra* note 30, at 200–02.

178 *See* Loewenstein, *supra* note 119, at 645 (listing insurrection attempts in the twentieth century and drawing the conclusion that "fascist strategists have grown particularly careful not to commit any overt act of rebellion until the subtler and studiously lawful methods of undermining the state and establishing the atmosphere of double legality warrant the ultimate seizure of power by *coup d'état*").

179 *See* Trump v. Anderson, 601 U.S. 100, 109–10, 115 (2024).

180 *Id.* at 123 (Sotomayor, J., concurring).

181 *Id.* at 122.

media would reveal "how hollow . . . the victories of the freedom of speech" may be.[182] A primary purpose of the First Amendment, at least according to Meiklejohn and Justice Brandeis,[183] is to serve democracy. This does not mean, of course, that only purely democratic views can be expressed. Its purpose is to allow the widest possible expression of diverse views. But what if this marketplace becomes distorted? What if the different views are not represented equally or fairly, but instead are influenced by powerful individuals?[184] And what if this development threatens democracy itself?

It has long been established that the Bill of Rights is not a suicide pact. This idea goes back in part to Thomas Jefferson, who expressed that the importance of saving the country "when in danger . . . [was the] higher obligation."[185] Later, Abraham Lincoln, the great Reconstructionist, expressed his understanding of certain exceptions in extreme cases when justifying the disregard of habeas corpus provisions in the Constitution.[186]

Finally, in his dissent in *Terminiello v. City of Chicago*, Justice Robert H. Jackson coined the phrase that "the constitutional Bill of Rights . . . [is not] a suicide pact."[187] It would later be echoed by many, including Justice Arthur Joseph Goldberg[188] and American legal philosopher, jurist, and scholar Ronald Dworkin.[189] This idea has always been closely associated with foreign affairs, especially in the aftermath of the 9/11 terror attacks, after which it has been used almost exclusively in the context of traditional security concerns.[190] It was within this

182 MEIKLEJOHN, *supra* note 60, at 87.

183 *See supra* Section II.A.

184 Meiklejohn also believed that while the suppression of freedoms is generally impermissible, the suppression of abuses of freedoms is generally permissible. *See* Alexander Meiklejohn, *What Does the First Amendment Mean?*, 20 U. CHI. L. REV. 461, 474 (1953).

185 5 THOMAS JEFFERSON, THE WRITINGS OF THOMAS JEFFERSON 542 (H.A. Washington ed., 2016) (ebook).

186 *See* 18 AM. HIST. LEAFLETS, LINCOLN'S INAUGURAL AND FIRST MESSAGE TO CONGRESS, 1861, at 16–18 (Albert Bushnell Hart & Edward Channing eds., 1912).

187 Terminiello v. Chicago, 337 U.S. 1, 37 (1949) (Jackson, J., dissenting).

188 *See* Kennedy v. Mendoza-Martinez, 372 U.S. 144, 160 (1963) ("[W]hile the Constitution protects against invasions of individual rights, it is not a suicide pact.").

189 *See* Ronald Dworkin, *The Threat to Patriotism*, N.Y. REV (Feb. 28, 2002), https://www.nybooks.com/articles/2002/02/28/the-threat-to-patriotism/ [https://perma.cc/LRZ4-CVQB].

190 *See id.*; *see also* RICHARD A. POSNER, NOT A SUICIDE PACT: THE CONSTITUTION IN A TIME OF NATIONAL EMERGENCY 1–3 (2006).

context when American legal scholar, Richard A. Posner, argued that even content-based restrictions could be constitutional.[191]

While the justification for restricting constitutional protections seems to be strongly tied to enemies from outside of the United States, there appears to be no good reason why it should not apply internally as well. The idea that the Bill of Rights should not contribute to its destruction remains the same. Abraham Lincoln himself recognized that the greatest threat to American democracy comes from within.[192] Lincoln's expression—"[A]re all the laws but one to go unexecuted, and the government itself go to pieces lest that one be violated?"[193]—is similar to Loewenstein's statement: "If democracy believes in the superiority of its absolute values over the opportunistic platitudes of fascism, it must live up to the demands of the hour, and every possible effort must be made to rescue it, even at the risk and cost of violating fundamental principles."[194] Both believed that when the survival of democracy is at stake, certain civil rights can be restricted.[195]

Some fear that embracing this principle will lead to a different hierarchy of values.[196] But such a hierarchy is only logical if one truly believes in the sovereignty of the people, as expressed in the first three words of the Preamble to the U.S. Constitution.[197] If "We the People" is to be taken seriously, the preservation of this sovereignty as the only means to self-government must have a special value.[198]

c. What About the Founding Fathers?

One could argue that the text of the U.S. Constitution sought to find the perfect balance between defensiveness and democratic liberty and that the Founding Fathers wanted to limit the defensive aspects to those exclusively mentioned in the text of

---

191 See *id*. at 125.

192 Abraham Lincoln, *The Perpetuation of Our Political Institutions (Address by Abraham Lincoln Before the Young Men's Lyceum of Springfield, January 27, 1838)*, 6 J. ABRAHAM LINCOLN ASS'N 6, 7 (1984).

193 AM. HIST. LEAFLETS, *supra* note 186, at 18.

194 Loewenstein, *supra* note 92, at 432.

195 On the similarities between the philosophies of Loewenstein and Lincoln, see MARK CHOU, DEMOCRACY AGAINST ITSELF: SUSTAINING AN UNSUSTAINABLE IDEA 70 (2014).

196 *See, e.g.*, Krotoszynski, *supra* note 129, at 1552.

197 MEIKLEJOHN, *supra* note 60, at 18.

198 *Id.* at 18–19 (explaining this logical hierarchy and that "[t]o that fundamental enactment [of self-government] all other provisions of the Constitution . . . are subsidiary and dependent").

the Constitution. Interpreting an overarching principle of militancy to affect other non-militant institutions would be disrespectful to the will of the Founders. On a merely textual basis, this argument is persuasive. But it fails to consider the two established functions of the First Amendment. Even though one main purpose of freedom of speech is to promote individual liberty, it also plays a vital role in preserving democracy.[199] The exact balance of these two purposes is certainly up to debate.[200] But it should be undisputed that there must be some sort of balance. However, when the liberal function of the First Amendment doctrine becomes so dominant that it completely undermines its function as a "cornerstone" of democratic self-government, then it has lost its balance.[201]

One could even argue that such an imbalance ignores the will of the Founders. Many of the Founding Fathers repeatedly expressed concern about the fragility of democracy. Despite the significant threats to the young democracy from the outside, the Founding Fathers already seemed to be more anxious about the threat from within. Thomas Jefferson worried that American democracy might degenerate into a form of "elective despotism."[202] Fearing the same, Alexander Hamilton recognized that the presiding president possessed an unparalleled ability to create disunity among people.[203] James Madison was concerned about the power of too mighty factions in democracy.[204] Similarly, George Washington's disdain of parties was fueled by his fear of polarization.[205] And Benjamin Franklin's famous statement—"A republic, if you can keep it"—can be interpreted as expressing Franklin's recognition that sometimes "the people" themselves are the greatest threat to their self-government.[206]

These expressions of concern by the Founders about the self-destructive tendencies of democracies, as well as constitutional history, and to some extent constitutional doctrine, point to the conclusion that U.S. constitutional law

---

199 *See supra* Section II.A.

200 *See supra* Section II.A.

201 Meiklejohn, *supra* note 60, at 55–56.

202 THOMAS JEFFERSON, NOTES ON THE STATE OF VIRGINIA 128–29 (1853).

203 *See* THE FEDERALIST NO. 68, at 342–43 (Alexander Hamilton) (Gideon ed., 2001).

204 *See* THE FEDERALIST NO. 10, *supra* note 158.

205 *See* George Washington, *Farewell Address (Sept. 17, 1796)*, *in* GEORGE WASHINGTON PAPERS, SERIES 2, LETTERBOOKS 1754–1799.

206 *See September 17, 1787: A Republic, if You Can Keep It*, NAT'L PARK SERV., https://www.nps.gov/articles/000/constitutionalconvention-september17.htm [https://perma.cc/LT79-LX9A] (Sept. 22, 2023).

could be susceptible to a militant democracy-type argument. Of course, some parts of the text, history, and doctrine are ambiguous. And yet there is a strong body of evidence showing that American constitutional history is marked by fears for its future and the survival of its democracy.

2. Danger of Militant Democracy and Populism

Even if constitutional doctrine is receptive toward militant democracy, there are political arguments to be made against its deployment. Some contend that militant democracy threatens democracy more than it secures it by providing authoritarian leaders with the very tools they need to abolish certain fundamental freedoms.[207] This is undoubtedly true. Still, it oversimplifies the problem. Of course, militant democracy has its inherent risks, but the failure to act may in many cases pose a risk just as great. Also, the dual nature of militant democracy is intended to minimize these inherent risks. As discussed above, on the one hand, preventive militant democracy seeks to prevent the enemies of democracy from coming to power. Thus, when applied consciously and as proposed in this Article, it aims to create a democracy-friendly environment that fosters a functioning and inclusive space for public discourse.[208] On the other hand, Institutional militant democracy places a check on this power by ensuring the independence of institutions, such as the courts. The combination of these two principles results in a self-limiting institutional democratic resilience that is backed by a democratic population and is therefore less vulnerable to abuse. Nevertheless, militant democracy measures must always be used with caution because of the inherent risk of abuse.[209]

One might ask how *Militant Democracy* addresses one of the biggest issues raised by social media: populism. Populists are not necessarily overt enemies of constitutional democracy, and they "do not come to power unfairly."[210] The threat to democracy posed by populism is more subtle. The emotional populist narrative undermines rational public debate under the guise of being the

---

207 *See* Carlo Invernizzi Accetti & Ian Zuckerman, *What's Wrong with Militant Democracy?*, 65 POL. STUD. 182, 183 (2017).

208 *See* Wu, *supra* note 21, at 568 (arguing for protecting and promoting a "healthy political speech environment").

209 *See* Stone, *supra* note 121, at 38–39. This view also prevails in Germany. *See* VOLKMANN, *supra* note 139, at 281 (arguing that too much defensiveness "must be prevented at all costs in order to protect the democratic process").

210 Sajó, *supra* note 5, at 194.

"voice of the people" and enforcing the true "will of the people."[211] But democracy is built on rational public discourse,[212] which is essential to any self-government.[213] Loewenstein's fear of emotionalism is even more relevant in today's world, where social media makes emotional manipulation considerably easier. He was convinced that "[p]reventing dangerous emotionalism and the consolidation of anti-liberal, authoritarian and substantively anti-democratic attitudes is a major task for those who would like to protect constitutional democracy."[214]

The populist motivation to enforce the true will of the people often leads to the diminution of minority rights[215] and the strengthening of centralized power,[216] which in turn is more prone to abuse of power and ultimately to despotism. At the same time, the emotional rhetoric dilutes democracy, establishing a superficial democracy. Additionally, the success of populists is usually based on manipulation instead of carefully considered political decisions. This can hardly be called real self-government of the people.[217]

These frequently voiced criticisms, while justified, should not stand in the way of considering militant democracy-influenced measures. They can be an effective tool against the increasing emotionalization of public discourse by populism. Furthermore, their misuse poses a risk that is substantially smaller than the risk posed by rising populism.[218]

3. Implications

It has been established that U.S. constitutional law could be sympathetic to militant democracy. But what follows from this

---

211 On this vox popoli narrative with a focus on the German AfD, see Oliver Schmidtke, *The 'Will of the People': The Populist Challenge to Democracy in the Name of Popular Sovereignty*, 32 SOC. & LEGAL STUD. 911, 912, 916, 922, 926 (2023).

212 Lincoln, *supra* note 192, at 13 ("Reason, cold, calculating, unimpassioned reason, must furnish all the materials for our future support and defence."). On the role of the Establishment Clause in keeping politics rational, see Sajó, *supra* note 92, at 567.

213 On the differences and difficulties of self-government compared to alien government, and on the importance of the former in American culture and history, see MEIKLEJOHN, *supra* note 60, at 11–12.

214 Sajó, *supra* note 5, at 201.

215 *The Dangerous Rise of Populism*, HUM. RTS. WATCH, https://www.hrw.org/world-report/2017/country-chapters/global-4 [https://perma.cc/K3UB-L8T6] (last visited Apr. 26, 2025); *see also* TAKIS S. PAPPAS, POPULISM AND LIBERAL DEMOCRACY 189, 204 (2019).

216 *See* PAPPAS, *supra* note 215, at 190–204.

217 On the legitimacy problems of a populist government, see Sajó, *supra* note 5, at 189–92.

218 *See* Huq, *supra* note 112, at 1141 (arguing that this "highly imperfect approach . . . may well be better than anything now available in light of the failure to act earlier").

recognition? It is again important to note that this Article does not claim to, nor does it intend to, solve all the problems of the online marketplace of ideas. Rather, it simply hopes to bring the argumentative figure of militant democracy to the center of the First Amendment debate. How and to what extent this figure can and will be used, this Article does not presume to judge. However, there are a vast number of possible applications of militant democracy regarding social media.

For one, militant democracy could be used strategically to enable the government to regulate social media[219] and promote independent, reliable media.[220] It could help to oblige platforms to change their algorithms to work in a more balanced and less emotional way, or to label or remove false information,[221] and, in general, to change the role and accountability of social media platforms, which bear a huge responsibility in this crisis of weakening democratice principles.[222] The European Democracy Action Plan could serve as an example in this regard.[223]

Additionally, it may help to introduce a "fairness doctrine" into the digital discourse to promote a balanced presentation of views.[224] On the issue of terrorist propaganda, militant democracy could be a viable argument to loosen the imminence requirement of *Brandenburg*'s imminent danger test.[225] And it might find an application in any upcoming proceedings regarding the TikTok ban.

---

219 *See* Jörg Ukrow, *Wehrhafte Demokratie 4.0—Grundwerte, Grundrechte und Social Media-Exzesse*, 24 ZEITSCHRIFT FÜR EUR. STUD. 65, 75 (2021).

220 On the importance of public service broadcasting, see *Recommendation No. R (96) 10 of the Committee of Ministers to Member States on the Guarantee of the Independence of Public Service Broadcasting*, COUNCIL OF EUR. 50–51 (Sept. 11, 1996), https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=090000168050c770 [https://perma.cc/8VT9-8XXA]; *see also* Sajó, *supra* note 5, at 202.

221 *See* Murthy v. Missouri, 603 U.S. 43, 50 (2024); Cass R. Sunstein, *A Framework for Regulating Falsehoods*, *in* SOCIAL MEDIA, FREEDOM OF SPEECH AND THE FUTURE OF OUR DEMOCRACY 53, 58, 61–62 (Lee C. Bollinger & Geoffrey R. Stone eds., 2022) (discussing the regulation of falsehoods and introducing a possible framework).

222 Netanel, *supra* note 22, at 510 ("[B]y propagating disinformation, platforms' recommender systems and the third-party tools that exploit them greatly magnify the force of democracy-destabilizing speech.").

223 *See id.* at 557 (applying Militant Democracy to social media with a focus on the European Democracy Action Plan).

224 SUNSTEIN, *supra* note 30, at 84–85 (arguing for a fairness doctrine in the digital realm).

225 *See* TYULKINA, *supra* note 92 (arguing that Militant Democracy could give anti-terrorism measures more legitimacy).

IV. BACK TO THE MARKETPLACE

The First Amendment is "the cornerstone of the structure of self-government."[226] But social media is no longer the marketplace of ideas that Justice Holmes or the Founders envisioned. It is broken; it favors the powerful and is used as a tool to manipulate the people. Misinformation does not contribute qualitatively to a free market of ideas.[227] Echo chambers prevent counterspeech on a significant scale. The culmination of these factors poses a major threat to democracy.

The current First Amendment doctrine is rightfully based on the idea that the political public sphere and public discourse are "the rock on which our government stands."[228] But that sphere and discourse must be inclusive, rational, and reliable to enable citizens to make their own considered judgments and form fact-based political views.

The concept of militant democracy is partly reflected in the U.S. Constitution and can be used both as an argument for a change in First Amendment doctrine and to improve the digital communicative sphere in general. This could be a first step in reviving the real marketplace of ideas in the digital realm.

---

226 MEIKLEJOHN, *supra* note 60, at 55.

227 When it comes to misinformation, the marketplace tends to fail. Instead of furthering truth, its dynamics tend to increase the credibility of falsehoods. *See* Cass R. Sunstein, *Falsehoods and the First Amendment*, 33 HARV. J.L. & TECH. 387, 393–95 (2020).

228 MEIKLEJOHN, *supra* note 60, at 77.

Technological Change and the Evolution of Patent Rights

Paul Rogerson

Technological Change and the Evolution of Patent Rights

Paul Rogerson∗

The history of patent law is characterized by a striking pattern: dramatic, recurring cycles between eras of stronger and weaker patent rights. One theory is that patent law has been adapting to waves of technological change. In periods of rapid technological progress, following major breakthroughs (like the steam engine or the microprocessor), freely granting patents tends to create thickets of overlapping claims that block innovation, and courts have reacted by weakening rights (raising the standard to obtain and enforce a patent).

Existing histories, relying on qualitative methods, have argued that this theory explains certain nineteenth-century legal developments, but have questioned whether it can explain the post-nineteenth-century period. This Article introduces new, complementary evidence that is based on a quantitative approach and covers substantially the entire two-century doctrinal history.

In particular, the Article presents a model for the correlation predicted by the technological theory (changes in the law lag behind changes in the pace of technological progress) using a dynamic ordered probit. To measure the rate of technological progress, it uses two quantitative measures from economic history: (1) the rate of productivity growth, and (2) the number of technologically significant patents issued by the United States Patent and Trademark Office. To measure changes in the law over time, it focuses on changes to an especially central doctrine—the standard of invention—and uses the judgments of doctrinal commentaries to code the law in each year.

The results both replicate the prior qualitative analyses of the nineteenth century and show that the same story extends to future periods. For nearly two centuries, courts have consistently weakened patent rights after the rate of technological progress rose, and vice versa, at a lag of roughly a decade. A one-percent rise in the rate of productivity growth, or one more technologically significant patent in force per five thousand people, has been associated with at least one step upward in the stringency of the standard of invention.

∗ Visiting Assistant Professor, Chicago-Kent College of Law. For helpful comments, the author thanks Kathy Baker, Sarah Burstein, Graeme Dinwoodie, Jonathan Masur, Greg Reilly, Mark Rosen, Dave Schwartz, Noah Smith-Drelich, and James Tierney, as well as audiences at IPSC, WIPIP, and PatCon.

I. INTRODUCTION

A major debate in patent law is whether technological progress will be best encouraged by "strong" patent rights (lenient legal standards that make patents easy to obtain and enforce) or "weak" patent rights (strict legal standards that make patents difficult to obtain and enforce).[1] This choice is thought to involve a tradeoff between the benefits of privatization and the benefits of open access.[2] On the one hand, stronger, more expansive patent rights may encourage innovation by creating more powerful incentives for inventors. On the other hand, they may also discourage innovation by fencing off the intellectual commons and blocking the work of future inventors.

The record of changes in the strength of patent rights over more than two hundred years since the founding of the U.S. patent system reveals a striking feature: dramatic, recurring cycles between eras of stronger and weaker patent rights. The most recent of these cycles is well known.[3] In the 1980s, courts and legislators began to worry that excessively weak patent protection in the middle decades of the twentieth century had discouraged innovation, so they moved to strengthen patent rights in a variety of ways.[4] The mood then swung in the opposite direction in the 1990s and early 2000s; a new generation worried that the reforms of the 1980s had spawned a "wave of bad patents" that would choke off innovation, and these reformers weakened patent rights in the 2000s and 2010s.[5]

But histories of patent law report that these cycles are nothing new. Instead, they seem to repeat continuously throughout the history of patent law, tracing back to its earliest days. For example, more than a century ago, one could find the leading patent law treatise of the late nineteenth century complaining bitterly (in words that might have been written today) of the "remarkable changes in the attitudes of our courts towards

---

1 *E.g.*, Mark A. Lemley, *The Surprising Resilience of the Patent System*, 95 TEX. L. REV. 1, 50 (2016) ("Much of the academic and policy debate over patent law in the past twenty years has focused on the relative dangers of overprotection and underprotection.").

2 *See e.g.*, Petra Moser, *Patents and Innovation: Evidence from Economic History*, 27 J. ECON. PERSPECTIVES 23, 23 (2013) ("[P]roviding stronger patents for early generations of inventors may also weaken incentives to invest in research and development for later generations . . . ."); Suzanne Scotchmer, *Standing on the Shoulders of Giants: Cumulative Research and the Patent Law*, 5 J. ECON. PERSPECTIVES 29, 30, 33 (1991); Saul Levmore, *Property's Uneasy Path and Expanding Future*, 70 U. CHI. L. REV. 181, 184–85 (2003).

3 *See* Lemley, *supra* note 1, at 7–13.

4 *Id.*; *see also* Ferring B.V. v. Barr Lab'ys, Inc., 437 F.3d 1181, 1195 (Fed. Cir. 2006) (Newman, J., dissenting) ("[P]atents were not a reliable support for commercial investment.").

5 Lemley, *supra* note 1, at 8–11.

patentees during the past few years," swinging from "an extreme liberality" in the early decades of the nineteenth century to "an equally extreme strictness" at the century's close.[6] This dynamic has been given a long list of nicknames by doctrinal commenters, such as the "swinging pendulum in patent law,"[7] "recurrent fluctuations" in patent rights,[8] and "cyclical judicial hostility toward patents."[9]

Legal scholars have offered three different explanations of the source of this dynamic, each with different lessons for patent law.[10] One story is that patent law has been trapped in a pathological cycle of overcorrection. For example, Professor Mark A. Lemley suggests that "[t]hese may in fact be cycles of over- and underprotection," where "every generation of Congresses, judges, and patent lawyers sees the imperfections of the decades before and overreacts to it."[11] Likewise, economists Adam B. Jaffe and Josh Lerner argue in their history of debates over patent rights that "[f]estering problems often lead to a dramatic shift in patent administration, which often goes too far in the other direction, generating a new set of problems of its own," which in turn sows the seeds of the next overcorrection.[12] In this story, the lesson of history is that legal actors should resist calls for dramatic reform of the patent system and instead focus on incremental changes.

A second story is that swings in patent law have been driven by political competition between interest groups.[13] Patent law not only creates incentives for innovation *ex ante*—it also divides the spoils *ex post*. Different legal regimes distribute wealth in distinct

---

6 1 WILLIAM C. ROBINSON, THE LAW OF PATENTS FOR USEFUL INVENTIONS 36, § 23 n.1 (1890).

7 Lemley, *supra* note 1, at 13; *see* ADAM B. JAFFE & JOSH LERNER, INNOVATION AND ITS DISCONTENTS 79 (2007).

8 H.R. Mayers, *The United States Patent System in Historical Perspective*, 3 PAT., TRADEMARK & COPYRIGHT J. RSCH. & EDUC. 33, 50 (1959).

9 Simone A. Rose, *Patent "Monopolyphobia": A Means of Extinguishing the Fountainhead?*, 49 CASE W. L. REV. 509, 526–27 (1999).

10 Readers will notice that each of these stories bears at least a family resemblance to a classic story that has been told about the evolution of *real* property rights. *See, e.g.*, Carol M. Rose, *Crystals and Mud in Property Law*, 40 STAN. L. REV. 577, 595–604 (1988) (property law's internal dynamics produce cycles between competing approaches); Harold Demsetz, *Toward a Theory of Property Rights*, 57 AM. ECON. REV. 347, 350–54 (1967) (changes in property rights have been a rational adaptation to changing social conditions); Saul Levmore, *Two Stories About the Evolution of Property Rights*, 31 J. LEGAL STUD. S421, S427 (2002) (changes in property rights have been the product of interest-group politics).

11 Lemley, *supra* note 1, at 14.

12 JAFFE & LERNER, *supra* note 7.

13 *See, e.g.*, STEPHEN H. HABER & NAOMI R. LAMOREAUX, THE BATTLE OVER PATENTS: HISTORY AND POLITICS OF INNOVATION 1–5 (2021); Eli Dourado & Alex Tabarrok, *Public Choice Perspectives on Intellectual Property*, 163 PUB. CHOICE 129, 146–47 (2015).

ways, and over time firms with differing business models have lobbied for their preferred outcomes, with varying success. The content of patent law in any given era has therefore been "the product of contending interests battling over surplus . . . ."[14] And the lesson of history is that would-be patent reformers should be skeptical of the pretexts for patent reform offered by interested parties—these same stories have been told many times before, in service of the same economic interests.[15]

A third story is that patent law has been adapting to waves of technological change.[16] Historians of technology have often argued that technological progress tends not to occur at a smooth, constant rate over time, but rather in "leaps" or "waves," driven by the introduction of major new technologies (like the steam engine or the microprocessor) that temporarily open up new opportunities for follow-on innovation, sometimes called "macroinventions" or "general purpose technologies."[17] The theory is that, during these technological booms, strong patent rights tend to hinder innovation—freely granting patents tends to clog the intellectual commons with conflicting legal claims. And, when these technological waves have occurred, courts have seen legal conflicts in litigated patent cases as well as public complaints from inventors, developed negative views of patents (without necessarily holding a worked-out theory about *why* strong patent rights seem to be creating problems), and reacted by weakening patent rights. Eventually, the wave of progress abates, the conflicts wane, and a new generation of judges come to view patents more positively.

The evidence for this third, technological theory is based largely on historical studies of the nineteenth century, which focus on the effects of the Second Industrial Revolution on an especially central doctrine of patent law—the standard of invention.[18] This legal doctrine sets the bar for how original an idea must be to

---

14 HABER & LAMOREAUX, *supra* note 13, at 4.

15 *See id.*

16 *See, e.g.*, William K. Townsend, *Patents, 1701–1901*, *in* TWO CENTURIES' GROWTH OF AMERICAN LAW, 1701–1901, at 394–96 (Members of the Faculty of the Yale Law Sch. eds., 1901); Edmund W. Kitch, Graham v. John Deere Co.*: New Standards for Patents*, 1966 SUP. CT. REV. 293, 319–22; B. Zorina Khan, *Technological Innovations and Endogenous Changes in U.S. Legal Institutions, 1790–1920*, at 16–17 (Nat'l Bureau of Econ. Rsch., Working Paper No. 10346, 2004); John F. Duffy, *Inventing Invention: A Case Study of Legal Innovation*, 86 TEX. L. REV. 1, 17–18 (2007).

17 *See, e.g.*, Joel Mokyr, *Punctuated Equilibria and Technological Progress*, 80 AM. ECON. REV. 350, 350–52 (1990); Timothy F. Bresnahan & M. Trajtenberg, *General Purpose Technologies 'Engines of Growth'?*, 65 J. ECONOMETRICS 83, 83–84 (1995).

18 *See* sources cited *supra* note 16.

qualify for patent protection. A lower standard demands less originality (allows more patents and protects "stronger" rights), while a higher standard demands more originality (allows fewer patents and protects "weaker" rights). In the early decades of the nineteenth century, patent law set a very low standard, and then raised it following the Civil War.[19] The theory is that this rise in the standard of invention was a response to the faster pace of technological progress ushered in by the Second Industrial Revolution in the late nineteenth century. This "quickening pace of technological progress . . . threatened to bring every commodity within a private patent grant," and courts reacted by tightening patentability standards.[20]

The analysis in these histories is extremely plausible, but it is also limited in some respects. One is that the arguments rely on a qualitative, informal notion of technological progress, which raises questions about how closely the legal and technological timelines correspond. For example, Townsend's history provides a list of "great inventions" that supposedly generated the legal conflicts in the late nineteenth century.[21] However, a number of these inventions were either developed early in the nineteenth century, decades before the law switched to a higher standard,[22] or had their most consequential social impacts only decades afterward, in the twentieth century.[23]

A second limitation is scope. These histories are largely focused on events in the nineteenth century. But the story does not end there—the standard of invention has continued to cycle up and down throughout the twentieth and twenty-first centuries. And, to the extent these histories have considered later time periods, they have been skeptical that legal changes in the twentieth and twenty-first centuries are explained by technological forces, and have instead viewed them as mistakes or confusions by the courts.[24] But it is difficult to evaluate these claims without a clear measure of how the pace of progress has changed over time.

The goal of this Article is to add new, complementary evidence on the technological theory. The most basic prediction of this

---

19 *See* Kitch, *supra* note 16, at 309–19.

20 *Id.* at 304.

21 *See* Townsend, *supra* note 16, at 395.

22 *See, e.g.*, O'Reilly v. Morse, 56 U.S. 62, 107–12 (1853) (telegraph).

23 *See, e.g.*, Paul David, *The Dynamo and the Computer: An Historical Perspective on the Modern Productivity Paradox*, 80 AM. ECON. REV. 355, 356–59 (1990) (electricity).

24 *See* Kitch, *supra* note 16, at 323–24 (describing later changes to the standard of invention as a "story of failure"); Duffy, *supra* note 16, at 41, 67 (arguing that twentieth-century legal changes might be best viewed as unsuccessful "experiments").

theory is simply that legal and technological changes have been correlated over time—the law has lagged behind movements in the pace of technological progress. The Article presents a new test of whether this prediction matches the broad shape of the doctrinal history. Its strategy has two main features.

One, the Article continues to focus on changes to the standard of invention. This doctrine is a good choice not only because existing histories have started there, but also because it has several desirable features that make it a useful and revealing window into patent law as a whole. First, it is often considered the single most important component of the strength of patent rights. Townsend's history reports that, in the nineteenth century, "[t]he changes in the law on this subject make the history of patent law in the United States."[25] And the doctrine has kept its central place in the legal scheme. Modern commenters have called today's standard of invention "the most important of the basic patent requirements," "the key to defining what is a patentable invention," and "the ultimate condition of patentability," among other superlatives.[26] In modern patent litigation, this doctrine is "the most common invalidity issue" and is "contested in a large majority of patent cases."[27] Second, historically, the standard of invention has also been the most consistent policy lever used to change the strength of patent rights. Every era with a substantial change in the legal system's basic attitude toward patents has seen that change embodied in a revision to the standard of invention (alongside a changing mix of revisions to various other doctrines).[28] Third, and finally, the evolution of this doctrine is (relatively) straightforward to describe. Legal commentaries have described the law as cycling between three major legal regimes that impose higher or lower standards of patentability: (1) "substantial novelty" (lowest standard); (2) "nonobviousness" (intermediate standard); and (3) "genius" (highest standard).[29]

Two, with this focus, the Article uses a quantitative approach. In particular, it presents a model for the suggested historical relationship between the pace of technological progress and the

---

25 Townsend, *supra* note 16, at 396.

26 Michael Abramowicz & John F. Duffy, *The Inducement Standard of Patentability*, 120 YALE L.J. 1590, 1593 (2011).

27 Dmitry Karshtedt, *Nonobviousness: Before and After*, 106 IOWA L. REV. 1609, 1611 & n.3 (2021).

28 *See infra* Part II.

29 *See* Kitch, *supra* note 16, at 297–303. Notice that Kitch uses the label "novelty" rather than "substantial novelty." *See id.*; Duffy, *supra* note 16, at 18.

standard of invention (changes in the law lag behind changes in technology), using a dynamic ordered probit framework.[30] This framework is able to capture two essential features of the legal history: (1) time dependence (the legal standard today is not independent of the legal standard yesterday), and (2) the nature of the available observations (legal regimes that can be ordered by *relative* strictness).

This Article's approach to data consists of the following. To measure the rate of technological progress in each year, the Article uses two major quantitative measures of technological change that have been developed by economic historians: (1) the rate of productivity growth, and (2) the number of technologically significant patents issued by the United States Patent and Trademark Office (USPTO). Both measures are available over the long run, extending back to the early or mid-nineteenth century, and therefore cover substantially all of the relevant doctrinal history. To measure changes in the standard of invention over time, the Article draws on the legal regimes used by doctrinal commentators to describe the evolution of the doctrine ("substantial novelty," "nonobviousness," and "genius"). In particular, relying on the judgments made by these commentaries, the Article codes the law in each year as falling into a particular legal regime.

The core result is a consistent, long-term correlation between the rate of technological progress and the standard of invention throughout substantially the entire two-hundred-year history of the U.S. patent system. This result both replicates the earlier analysis of the nineteenth century and shows that the same story extends to future periods. Qualitatively, the two series are well-matched—legal actors have switched to stricter legal regimes following rises in the rate of technological progress, and vice versa, at a lag of roughly a decade. And, in the model, the coefficient on the rate of technological progress is large, positive, and has a ninety-five percent confidence interval bounded above zero. Over the long run, a one-percent rise in the rate of productivity growth, or one additional technologically significant patent in force per five thousand people, is associated with at least one step upward in the strictness of the legal regime.

These results are broadly robust to a range of plausible changes in the model's inputs, including different estimates of the

30 *See* Barry Eichengreen et al., *Bank Rate Policy Under the Interwar Gold Standard: A Dynamic Probit Model*, 95 ECON. J. 725, 733–35 (1985).

rate of technological progress (from different economic historians) and different ways of coding the doctrinal history (reflecting the margins of disagreement among different historical commentaries), with one important qualification. The two measures of technological progress—productivity growth and technologically significant patents—track each other closely for most of the history but diverge to some extent in the most recent decades (1980s to early 2000s). Both measures show a recent wave of technological progress, but differ on its size. In these years, the productivity data is a somewhat better fit for the doctrine.

In sum, this Article makes two main contributions. First, it draws a missing link between patent law and the measurement of technological progress in economic history. It shows that the major quantitative measures of technological change used by economic historians track changes in the strength of patent rights over time. Second, it contributes to a long-running debate about the history of patent law. The results create clearer evidence that the doctrinal record fits the technological theory, providing a stronger reason to believe that cycles in patent law have been a *positive* feature of the legal history. Courts have plausibly adopted legal regimes that best encourage technological innovation in different social conditions.

The rest of the Article is organized as follows: Part II describes the doctrinal history, Part III explains the technological theory and presents a model that captures its core claim, Part IV describes the data, and Part V carries out the analysis. A brief conclusion follows.

II. DOCTRINAL HISTORY

This section begins by simply describing the changes in the standard of invention over time (which the technological theory seeks to explain). There is a broad consensus among commenters on the basic shape of this history. There are two noteworthy features worth highlighting at the top.

First, the courts have played a particularly large role in the development of this doctrine. The task of setting the standard of invention has always been delegated, in substantial part, to the courts. The Supreme Court has famously argued that this was always the plan—supposedly, Thomas Jefferson (head of the original patent board) felt that a standard of invention was necessary, but should be left to the courts to develop through case

law.[31] The veracity of this claim has been (convincingly) questioned,[32] but whatever the original intent, that is what happened. For the first century-and-a-half of the U.S. patent system, no statute set forth a general standard of invention; the doctrine was instead created by the courts through a common-law-like process.[33] The standard of invention was eventually codified by the Patent Act of 1952.[34] However, this legislation set forth a broad principle rather than detailed instructions, and, over time, different interpretations of the statute have waxed and waned in the case law.

Second, commenters have described the law as cycling between three basic legal regimes over time, which can be ranked as imposing higher or lower standards of invention. The regimes are as follows: (1) "substantial novelty" (lowest standard); (2) "nonobviousness" (intermediate standard); and (3) "genius" (highest standard).[35] Of course, it is not possible to make precise numerical statements about how strict or lenient each regime is, but it is at least possible to say that different regimes impose *relatively* higher or lower standards—that is, it is harder to obtain and enforce a patent under one regime compared to another. The cycles in patent law involve switches between these regimes over time (as well as occasional periods where the law was in between two regimes).

Broadly similar accounts of this history can be found in many different sources. Altogether, they describe six major historical periods from 1790 (the date of the first patent statute) to today. These six periods are discussed below.

A. "Substantial Novelty"

In the earliest decades of the patent system—the late eighteenth and early nineteenth centuries—patent law largely lacked a developed standard of invention, and was instead generally willing to grant a patent on any invention that was merely "novel" (literally new).[36] A famous early nineteenth-century

---

31 Graham v. John Deere Co., 383 U.S. 1, 5–10 (1966) ("[T]he whole was turned over to the judiciary, to be matured into a system . . . .").

32 Kenneth J. Burchfiel, *Revising the "Original" Patent Clause: Pseudohistory in Constitutional Construction*, 2 HARV. J.L. & TECH. 155, 218 (1989).

33 *See* Kitch, *supra* note 16, at 303–27.

34 *Id.* at 293.

35 *Id.* at 297–303; Duffy, *supra* note 16, at 18.

36 *See* Townsend, *supra* note 16, at 394–95.

case, *Earle v. Sawyer*, which was decided by Justice Story while riding circuit, emphatically states this doctrine:

> It is of no consequence whether the thing be simple or complicated – if it is new, if it is useful, if it has not been known or used before, it constitutes an invention within the very terms of the Act, and in my judgment within the very sense and intendment of the legislature.[37]

As a number of commenters have noted, the regime that prevailed in these early decades was not quite a pure "novelty" regime.[38] Instead, it has often been given the label "substantial novelty."[39] This "substantial novelty" regime prohibited patents not only on inventions that were literally identical to existing technology, but also on inventions that differed from existing technology only in trivial details, so that they could be said to be "substantially" identical.[40] A classic example was that merely varying the "form" or "proportions" of the parts of a machine—in a way that made no difference to the machine's function—would not be enough for a patent.[41] However, with that proviso, the law was otherwise inclined to issue patents.[42]

A testament to the leniency of this early standard is the fact that, unlike in modern patent litigation—where the standard of invention is "the most common invalidity issue"[43]—invention was rarely a central issue in patent litigation before the Civil War. Throughout almost the entire nineteenth century, the Supreme Court was the court of first (and last) appeal for patent disputes; it was not until the passage of the Evarts Act of 1891 that the intermediate circuits were created, and the Supreme Court was given certiorari power (at least in nonconstitutional cases).[44] During these years, the Supreme Court therefore carried a heavy patent docket. In the decades preceding the end of the Civil War, it issued eighty-three opinions in patent cases (with the lions'

---

37 Earle v. Sawyer, 8 F. Cas. 254, 256 (C.C.D. Mass. 1825) (No. 4247).

38 *See* Burchfiel, *supra* note 32, at 191–95; Edward C. Walterscheid, *The* Hotchkiss *Unobviousness Standard: Early Judicial Activism in the Patent Law*, 13 J. INTELL. PROP. L. 103, 106 (2005) [hereinafter Walterscheid, *Unobviousness Standard*]; Edward C. Walterscheid, *Novelty & the* Hotchkiss *Standard*, 20 FED. CIR. BAR J. 219, 227–28 (2010) [hereinafter Walterscheid, *Hotchkiss Standard*].

39 Walterscheid, *Hotchkiss Standard*, *supra* note 38, at 228.

40 *Id.* at 242–46.

41 *See* Burchfiel, *supra* note 32, at 196–97.

42 *Id.* at 191.

43 *See* Karshtedt, *supra* note 27, at 1611 n.3 (citing Apple Inc. v. Samsung Elecs. Co., 839 F. 3d 1034, 1074 (Fed. Cir. 2016) (en banc) (Dyk, J., dissenting)).

44 *See* DAVID P. CURRIE, THE CONSTITUTION IN THE SUPREME COURT: THE SECOND CENTURY, 1888–1986, at 5 (1990); *see also* Benjamin B. Johnson, *The Origins of Supreme Court Question Selection*, 122 COLUM. L. REV. 793, 823–28 (2022).

share coming in the 1840s, 1850s, and 1860s).[45] Yet, among these opinions, just a handful of them resolved a dispute over whether the standard of invention had been met.[46]

B. "Nonobviousness"

A new legal regime, "nonobviousness," then took hold in the post-Civil War years.[47] The essence of the "nonobviousness" regime is that "[not] all improvement is [] invention."[48] Instead, to be entitled to a patent, an invention "must involve something more than what is obvious to persons skilled in the art to which it relates."[49] In other words, even an invention that represents more than a "trivial" change to existing technology cannot be patented if it would have been an obvious idea to ordinary workers in the relevant technical field.[50]

The earliest signs of the "nonobviousness" test are located in the Supreme Court's first invention opinion, *Hotchkiss v. Greenwood*, a pre-Civil War case.[51] *Hotchkiss* did not use the word "obviousness," but it did refer to the skill of an "ordinary mechanic," which sounds at least obviousness-like, particularly to modern readers.[52] However, the common verdict, shared by all close histories of this period, is that this line in *Hotchkiss* was not understood as a doctrinal revolution at the time, "either by the Supreme Court or by legal commentators," and "substantial novelty" remained the governing test for the next several decades.[53] For example, the year after *Hotchkiss* was decided, the author of the opinion, Justice Nelson, gave instructions in a patent case (while riding circuit) that ignored *Hotchkiss*.[54] Another justice who had joined the opinion did the same in the following year.[55] The handful of Supreme Court invention cases during subsequent

45 This information comes from the Author's database of Supreme Court patent cases.

46 *See* Kitch, *supra* note 16, at 315–17 ("Cases before the Supreme Court raising issues of patentable novelty after *Hotchkiss* were infrequent.").

47 *See id.* at 319.

48 Pearce v. Mulford, 102 U.S. 112, 118 (1880).

49 *Id.*

50 Kitch, *supra* note 16, at 301–03.

51 Hotchkiss v. Greenwood, 52 U.S. 248 (1851); Walterscheid, *Unobviousness Standard, supra* note 38, at 104.

52 Walterscheid, *Unobviousness Standard, supra* note 38, at 104.

53 Burchfiel, *supra* note 32, at 202–09; Kitch, *supra* note 16, at 309, 314–16; Walterscheid, *Unobviousness Standard*, *supra* note 38, at 104; Walterscheid, *Hotchkiss Standard*, *supra* note 38, at 219–20.

54 Walterscheid, *Unobviousness Standard*, *supra* note 38, at 116 n.82 ("One of the more remarkable aspects of the early treatment of the *Hotchkiss* standard was that its author, Justice Nelson, one year later refused to follow it in a circuit court opinion.").

55 Burchfiel, *supra* note 32, at 205.

decades continued to apply the governing "substantial novelty" rules.[56] And *Hotchkiss* was not cited by the Supreme Court in any invention case until 1875 and not by any lower court until 1882.[57]

Instead, "[o]nly after the Civil War was non-obviousness regularly applied by courts as a general condition of patentability."[58] The establishment of "nonobviousness" as the prevailing test (and the initial working-out of this new doctrine) was an incremental process accomplished through a decade of case law. In these post-Civil War years, only "slowly does the non-obviousness test come to the forefront," and the first signs of the new regime were so subtle that they "may have been at the time imperceptible to both the Court and its bar."[59]

The beginning of the transition is generally dated to the early 1870s. In particular, two views about the timing are represented in the commentary. Edward C. Walterscheid and Kenneth J. Burchfiel identify 1870 as the starting point—that is the date of the Supreme Court's first post-Civil War invention case, *Stimpson v. Woodman*.[60] Edmund W. Kitch is more conservative, and argues that the invention cases of the early 1870s continued to apply the old "substantial novelty" rules.[61] He instead traces the first signs of the new "nonobviousness" approach to a pair of opinions issued in 1875, one of which (*Collar Co. v. Van Dusen*) is the Supreme Court's first post-Civil War citation to *Hotchkiss*.[62]

The endpoint of the transition is generally dated roughly a decade later, in the early 1880s. Again, one can find a range of views on the exact timing. Benjamin F. Lee, a late nineteenth-century lecturer in patent law at New York Law School, reports that the transition "culminated" in the "leading case" *Pearce v. Mulford*, which held that a patentable invention "must involve something more than what is obvious to persons skilled in the art

---

56 *Id.* at 206–07 ("For nineteen years after the *Hotchkiss* decision, the Supreme Court did not invalidate a single patent on the basis of obviousness. More importantly, in sustaining the validity of patents before it, the Court routinely failed to conduct any inquiry into the degree of skill required to derive an invention over the disclosure of the prior art. . . . Lower courts and commenters continued to consider the degree of skill required legally immaterial."); *see also* Kitch, *supra* note 16, at 309–16.

57 Kitch, *supra* note 16, at 319 & n.120.

58 Burchfiel, *supra* note 32, at 207–08.

59 Kitch, *supra* note 16, at 319.

60 Burchfiel, *supra* note 32, at 208 n.301; *see* Walterscheid, *Unobviousness Standard*, *supra* note 38, at 104; Stimpson v. Woodman, 77 U.S. 117 (1870).

61 Kitch, *supra* note 16, at 318.

62 *Id.* at 318–19; *see* Collar Co. v. Van Dusen, 90 U.S. 530 (1875).

to which it relates."[63] Frank D. Prager locates the final doctrinal settlement in the 1882 case *Loom Co. v. Higgins*, which played a crucial role in establishing the doctrine of "secondary considerations" (or "objective evidence") as a key part of the obviousness analysis.[64] 1882 is also the date of a telling anecdote, relayed by Burchfiel. In that year, the Supreme Court allowed a defendant to raise obviousness on appeal for the first time, when it had not been raised in the lower court.[65] In other words, in the early 1880s, there were at least some patent lawyers in the United States who had not previously understood they should raise obviousness, and now did. 1883 is the year of *Atlantic Works v. Brady*, often considered the "classic" invention opinion of the period.[66] Again, being the most conservative, Kitch asserts that the new rule only became clear in cases in 1885.[67]

Supreme Court opinions of the time disclose the concerns that motivated the new approach. The Justices worried that a low standard of invention had produced overabundant, conflicting patent claims that ultimately served to block innovation. Justice Bradley's 1883 *Atlantic Works* opinion is the leading statement of these concerns:

> To grant to a single party a monopoly of every slight advance made, except where the exercise of invention somewhat above ordinary mechanical or engineering skill is distinctly shown, is unjust in principle and injurious in its consequences.
>
> . . . Such an indiscriminate creation of exclusive privileges tends rather to obstruct than to stimulate invention. It creates a class of speculative schemers who make it their business to watch the advancing wave of improvement, and gather its foam in the form of patented monopolies, which enable them to lay a heavy tax upon the industry of the country, without contributing anything to the real advancement of the art. It embarrasses the honest pursuit of business with fears and apprehensions of concealed liens and unknown liabilities to lawsuits and vexatious accountings for profits made in good faith.[68]

---

63 Benjamin F. Lee, *What Constitutes a Patentable Subject Matter*, 3 COUNSELLOR 159, 164–65 (1894) (quoting Pearce v. Mulford, 102 U.S. 112, 118 (1880)).

64 *See* Frank D. Prager, *Standards of Patentable Invention from 1474 to 1952*, 20 U. CHI. L. REV. 69, 80 (1952) (citing Loom Co. v. Higgins, 105 U.S. 580 (1882)).

65 Burchfiel, *supra* note 32, at 208.

66 ALBERT H. WALKER, TEXT-BOOK OF THE LAW OF PATENTS FOR INVENTIONS 26 (5th ed. 1917) (citing Atlantic Works v. Brady, 107 U.S. 192 (1883)).

67 Kitch, *supra* note 16, at 319 ("Not until 1885, in *Thompson v. Boisselier*, did the Court take pains to point out that 'it is not enough that a thing shall be new, in the sense that in the shape or form in which it is produced it shall not have been before known, and that it shall be useful, but it must, under the Constitution and the statute, amount to an invention or discovery.'") (quoting Thompson v Boisselier, 114 U.S. 1, 11 (1885)).

68 Atlantic Works v. Brady, 107 U.S. 192, 200 (1883).

C. "Genius"

"Nonobviousness" remained the prevailing test through the early decades of the twentieth century.[69] During these years, the terminology was not yet entirely standardized, so the new test sometimes went by other names, such as "ingenuity."[70] And there were a variety of "minor" trends and fluctuations in the case law—in some years, courts seemed to apply the test in ways that were somewhat stricter, or somewhat more lenient.[71] But these fluctuations occurred within a broadly established framework, whose fundamental rules were not in dispute.

Then, several decades into the twentieth century, a "major" trend emerged toward a distinctly higher standard.[72] The beginning of this shift is generally dated to the years around 1930. Judge Learned Hand—a prominent circuit judge, and the leading judicial chronicler of the Supreme Court's evolving attitudes towards patents during this period[73]—wrote in the 1942 opinion *Picard v. United Aircraft Corp.* that "the Supreme Court . . . has for a decade or more shown an increasing disposition to raise the standard of originality necessary for a patent," amounting to "a pronounced new doctrinal trend."[74] A later Judge Hand opinion, *Lyon v. Bausch & Lomb Optical Co.*, similarly gives the starting point of this trend as "twenty, or perhaps, twenty-five" years before 1952 (so, between 1927 and 1932).[75] Prager reports that the shift "started" under Chief Justice Hughes,[76] who ascended to the chiefship in 1930.[77] And Robert P. Merges tells us that 1930 is the starting point of the Supreme Court's "most virulent anti-patent era."[78] However, commentary at the time also acknowledged that—as with many gradual judicial trends—it was "difficult to establish which exact case first showed the Supreme Court's new

69 *See* Prager, *supra* note 64, at 81; *see also* Duffy, *supra* note 16, at 41.

70 *See* Prager, *supra* note 64, at 80–81.

71 *Id.*; *see also* Duffy, *supra* note 16, at 41 (explaining that decisions during this period were not always "consistent," with some cases seemingly applying a "high" standard and others seemingly applying a "lax" standard, but "[m]ore often, at least through the end of the nineteenth century and the first few decades of the twentieth century, the Court used language quite similar to the modern standard").

72 *See* Prager, *supra* note 64, at 81–82.

73 Richard A. Posner, *The Hand Biography and the Question of Judicial Greatness*, 104 YALE L.J. 511, 521 (1994).

74 Picard v. United Aircraft Corp., 128 F.2d 632, 636 (2d Cir. 1942).

75 Lyon v. Bausch & Lomb Optical Co., 224 F.2d 530, 535 (2d Cir. 1955).

76 Prager, *supra* note 64, at 81.

77 Currie, *supra* note 44, at 205–07.

78 Robert P. Merges, *One Hundred Years of Solicitude: Intellectual Property Law, 1900-2000*, 88 CAL. L. REV. 2187, 2223 (2000).

attitude toward patents,"[79] and commentaries offered a range of estimates around the year 1930 for the precise starting point of this trend, from as early as 1925[80] to the late1930s.[81]

The culminating point of this trend was the Supreme Court's "infamous" 1941 decision in *Cuno Engineering Corp. v. Automatic Devices Corp.*[82] It was in *Cuno* that the new regime "reached full maturity"[83] and "became clear."[84] *Cuno* called for a "[s]trict application" of the test for invention.[85] It also contained the notorious line that a patentable invention "must reveal the flash of creative genius."[86] The Supreme Court itself never coined a name for this new, "stricter" version of the "nonobviousness" test, preferring to insist that it was merely following the same old rules. But, taking their cue from *Cuno*, doctrinal commenters have generally referred to it as the "genius" or "flash of genius" test.[87]

There was some dispute as to what legal test was actually contemplated by the phrase "flash of creative genius," and two different views were advanced.[88] The "subjective" view interpreted this phrase to require that an invention be produced by a particular kind of mental process—a literally instantaneous "flash" of creative insight—thereby precluding protection for inventions that had been created through step-by-step experimentation.[89] On the other hand, the "objective" view construed "flash of creative genius" as merely a colorful way of saying that the standard was high, and viewed the test as simply requiring "objective" genius.[90] To the extent a judgment can be made, the "objective" interpretation seems better supported by the case law. Even in the 1940s, during the height of anti-patent sentiment on the Court, one of the few patents that every member of the Court was willing to sustain (including the anti-patent ringleaders Justices Douglas, Black, and Stone) was a patent on

79 Allen M. Krass, *Test of Patentability Under the 1952 Patent Act*, 2 WAYNE L. REV. 130, 133 (1956).

80 Donald J. Libert, *Section 103 of the Patent Act and the Standard of Invention: Comments on* Lyon v. Bausch & Lomb Optical Co., 38 J. PAT. OFF. SOC'Y 304, 308–11 (1956).

81 Krass, *supra* note 79, at 132–33.

82 Kitch, *supra* note 16, at 323 (citing Cuno Eng'g Corp. v. Automatic Devices Corp., 314 U.S. 84 (1941)).

83 Krass, *supra* note 79.

84 Prager, *supra* note 64, at 81.

85 *Cuno Eng'g Corp.*, 314 U.S. at 92.

86 *Id.* at 91.

87 *See, e.g.*, Kitch, *supra* note 16, at 300–01; Duffy, *supra* note 16, at 18.

88 Duffy, *supra* note 16, at 18 (discussing the "subjective" and "objective" interpretations of the test of patentability).

89 *Id.*

90 *Id.*

an invention created through step-by-step experimentation.[91] But, in any event, "genius" was a high bar.[92] As Justice Jackson acidly wrote in a 1949 dissent, the Supreme Court had developed such a "strong passion" for striking down patents that "the only patent that is valid is one which this Court has not been able to get its hands on."[93]

Much like in the late nineteenth century, opinions of the time described the concerns underlying these changes to the standard of invention. The opinions harshly criticized perceived abuses in the patent system and characterized patents as dangerous monopolies that would impede innovation, if granted too freely. The *Cuno* opinion is the leading expression of these views, and quotes Justice Bradley's remarks in *Atlantic Works* with approval:

> Strict application of [the test for invention] is necessary lest in the constant demand for new appliances the heavy hand of tribute be laid on each slight technological advance in an art. The consequences of the alternative course were forcefully pointed out by Mr. Justice Bradley in *Atlantic Works v. Brady*: "Such an indiscriminate creation of exclusive privileges tends rather to obstruct than to stimulate invention."[94]

D. The Patent Act of 1952

The next era of invention doctrine was ushered in by the passage of the Patent Act of 1952 (1952 Act).[95] This legislation, for the first time, codified the standard of invention, and used the language of "obviousness" to do it. The statutory standard, which largely endures to this day, provided that a patent will not be granted on an invention that "would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains," and added that "[p]atentability shall not be negated by the manner in which the invention was made."[96] These provisions were codified at 35 U.S.C. § 103 and are often referred to as "Section 103."

After the passage of this legislation, a debate immediately broke out among the circuit courts (and scholars) about its effect

---

91 *See* Graver Tank & Mfg. Co. v. Linde Air Prods. Co., 336 U.S. 271, 274 (1949) ("[The inventors] collaborated for some six months in conducting a series of about 500 experiments in the course of which they compounded 75 different flux compositions. They finally produced the invention for which a patent was sought.").

92 Duffy, *supra* note 16, at 42.

93 Jungersen v. Ostby & Barton Co., 335 U.S. 560, 572 (1949) (Jackson, J., dissenting).

94 Cuno Eng'g Corp. v. Automatic Devices Corp., 314 U.S. 84, 92 (1941) (citation omitted) (quoting Atl. Works v. Brady, 107 U.S. 192, 200 (1883)).

95 Patent Act of 1952, Pub. L. No. 82-593, 66 Stat. 792 (codified as amended at 35 U.S.C. §§ 1–293).

96 *Id.* § 103, 66 Stat. at 798 (codified as amended at 35 U.S.C. § 103).

on the standard of invention.[97] One view was that the 1952 Act lowered the standard of invention, turning the clock back to 1930 and undoing the prior two decades of Supreme Court precedent.[98] Another view was that the 1952 Act largely codified the existing doctrine (and the standard of patentable invention) as it stood in 1951. Under this view, the statute perhaps rejected the *subjective* version of the "genius" test (by providing that "[p]atentability shall not be negated by the manner in which the invention was made"), but it left *objective* "genius," and the general level of patentable invention, unchanged.[99]

Both sides found support in the legislative history, and ultimately the circuits split nearly evenly. Six of the eleven circuits took the position that Congress did not intend to lower the standard of patentability, but rather sought to codify preexisting case law, while the other five circuits held that Congress had intended that "the rigid test of patentability as laid down by the recent Supreme Court cases be somewhat relaxed by Section 103."[100] There is some evidence that these doctrinal differences were not distinctions without a difference. The circuits' "conflicting interpretations" were correlated with a "wide disparity in the percentage of patents being held valid—from 8% in the Eighth Circuit to 82% in the Tenth Circuit."[101]

This dispute was never clearly resolved by the Supreme Court. The Court's first post-1952 invention opinion—*Graham v. John Deere Co.*, issued fourteen years after the passage of the 1952 Act—confusingly seemed to point in both directions.[102] Contemporary commentary described the mixed signals that the decision sent to patent lawyers:

> The *Graham v. Deere* opinion generated some euphoria. It was mild and respectful of the patent system. It was free of the extreme expressions and special requirements to be found in some earlier

---

97 *E.g.*, John M. Webb, *The Changing Standard of Patentable Invention: Confusion Compounded*, 55 MICH. L. REV. 985, 990–97 (1957); Hillel Marans, *Some Aspects of the Patent Act of 1952 as Interpreted by Recently Published Decisions*, 39 J. PAT. OFF. SOC'Y 177, 189–91 (1957); Gay Chin, *The Statutory Standard of Invention: Section 103 of the 1952 Patent Act*, 3 PAT. TRADEMARK & COPYRIGHT J. RSCH. & EDUC. 317, 318–28 (1959); William J. Keating, *The Supreme Court Interprets the Patent Statute: A Trilogy of Cases and Their Effect Today*, 72 DICK. L. REV. 244, 249–50, 257–67 (1967).

98 Chin, *supra* note 97, at 323.

99 *See id.* at 321–22.

100 *Id.* at 322; *see also, e.g.*, Keating, *supra* note 97, at 257–65.

101 Keating, *supra* note 97, at 250. Of course, litigation statistics like these must be viewed with caution—cases are neither randomly selected for litigation nor randomly assigned to different circuits. But these data provide at least some evidence that the doctrinal differences were practically meaningful.

102 *See* Graham v. John Deere Co., 383 U.S. 1 (1966).

> majority and concurring opinions of the Court. . . . Further, the emphasis on *Hotchkiss v. Greenwood* tended to support a view that the Court had returned to a relatively mild test of patentability perceived by some persons as part of the "good old days."
>
> But other language in the *Graham v. Deere* opinion pointed in a different direction. The Court emphasized the strict standard of unobviousness, and made unkind comments on the Patent Office performance. The Court stressed its view that the level of patentability was not lowered by Section 103. . . . These and other aspects of the opinion suggest that the Court was not casting aside all of the views expressed in recent opinions on "invention" and that it continued to view the standard (however expressed) as quite severe.[103]

In practice, the opinion seemed to apply a somewhat less rigorous standard than that found in pre-1952 cases. Yet, in the years after *Graham*, the Supreme Court issued several opinions that seemed to diverge from the *Graham* approach and instead return to aspects of the pre-1952 framework.[104] And, more generally, during these decades, the Supreme Court's interest in supervising the circuit courts' patent jurisprudence waned.[105] The Court rarely granted certiorari in patent cases, and seemed to give something less than its full attention to the cases it did take, issuing opinions that were "poorly reasoned and inconsistent."[106]

Absent useful guidance from the Supreme Court, the circuit courts were left to their own devices, and the split persisted, with some circuits applying a stricter regime, and others a distinctly more lenient regime. Robert L. Harmon, a then-practicing patent lawyer and later the author of an influential patent law treatise, described this period as follows:

> When this author broke into the business, and for many years thereafter, it was quite clear that there was no such thing as a valid

---

103 George E. Frost, *Does Section 103 Apply to "Combination" Patents?*, 5 APLA Q.J. 117, 117 (1977) (footnote omitted).

104 *See* Herbert H. Mintz, *The Standard of Patentability in the United States—Another Point of View*, 1977 DET. COLL. L. REV. 755, 785 (arguing that *Graham* and subsequent cases created a doctrinal "[m]orass").

105 *See* John F. Duffy, *The* Festo *Decision and the Return of the Supreme Court to the Bar of Patents*, 2002 SUP. CT. REV. 273, 275.

106 Jaffe & Lerner, *supra* note 7, at 99–100 ("[I]mportant differences persisted because the Supreme Court rarely heard patent-related cases. The justices were reluctant to devote their time to what they saw as banal commercial disputes. For instance, one of the few patent cases that was heard by the 'Supremes' during the 1970s involved a dispute over a system to flush cow manure from the floor of dairy barns. Who would get the duty of writing the opinion for this 'cow shit case' was a matter of considerable controversy—ultimately, it was assigned to Justice William Brennan because he had antagonized the Chief Justice with his acrid dissents in other cases. Perhaps not surprisingly, the decision that resulted in this case was poorly reasoned and inconsistent with the Supreme Court's own earlier rulings." (footnotes omitted)).

> patent in the Eighth Circuit, and the climate in the Ninth Circuit was not much more hospitable. In the Seventh Circuit, on the other hand, patent infringement could get a client into big trouble. Each of the other circuits occupied its own bad in the enforcement spectrum.[107]

Again, there are signs that these doctrinal differences were practically meaningful—empirical studies report that large differences in patent-validity decisions across the circuit courts persisted throughout the 1970s.[108]

E. The Federal Circuit's "Suggestion" Test

These divisions were ultimately resolved by Congress's 1982 creation of the U.S. Court of Appeals for the Federal Circuit, a specialized court of appeals for patent cases to replace the regional circuits.[109] Congress was motivated by growing fears in the 1970s that divisions between the circuits and weak patent rights had undermined incentives for innovation.[110] The new court was intended both to unify patent law by resolving disputes that the Supreme Court had left to linger, and to strengthen patent rights by selecting rules that would make patents easier to obtain and enforce.[111] Judge Newman (appointed to the Federal Circuit in 1984) described the feeling on the new court this way:

> Practitioners from an earlier era well recall the adverse impact on industrial innovation when patents were not a reliable support for commercial investment, based in part on the judicial belief that patents and their practice were seriously flawed. With invalidation of most of the patents that reached the courthouse, the contribution of a diminished patent incentive to the weakening of technology-based investment was a serious national concern, and the impact on the nation's commercial vigor was recognized by government as well as by the industrial and scientific communities. The formation of the Federal Circuit Court of Appeals was part of the nation's program to restore technology-based leadership with the aid of an effective patent system.[112]

The Federal Circuit's first opinion, *South Corp. v. United*

---

107 Robert L. Harmon, *Seven New Rules of Thumb: How the Federal Circuit Has Changed the Way Patent Lawyers Advise Clients*, 14 GEO. MASON U. L. REV. 573, 574 (1992).

108 *See* Matthew D. Henry & John L. Turner, *The Court of Appeals for the Federal Circuit's Impact on Patent Litigation*, 35 J. LEGAL STUD. 85, 99–103 (2006).

109 *See* Federal Court Improvements Act of 1982, Pub. L. No. 97-164, 96 Stat. 25 (codified as amended in scattered sections of 28 U.S.C.).

110 *See* Rochelle Cooper Dreyfuss, *The Federal Circuit: A Case Study in Specialized Courts*, 64 N.Y.U. L. REV. 1, 6–7 (1989).

111 *Id.*; *see also* Lemley, *supra* note 1, at 7 ("[T]he Federal Circuit was created in 1982 with a mission to strengthen patent rights.").

112 Ferring B.V. v. Barr Lab'ys, Inc., 437 F.3d 1181, 1195–96 (2006) (Newman, J., dissenting) (citation omitted).

States, adopted as precedent the body of opinions that had been issued by the Court of Customs and Patent Appeals (CCPA), the specialized court that previously heard appeals from the PTO (a duty taken over by the Federal Circuit).[113] The CCPA had always been a pro-patent venue, which arguably never followed *Cuno* in the first place, and certainly had not done so after the 1952 Act.[114] *South Corp.* therefore rejected a "strict" standard of invention in the Federal Circuit, a message reinforced by early opinions that were sharply critical of the Supreme Court's patent jurisprudence.[115]

The Federal Circuit then gradually developed a new rule of evidence—known as the "suggestion" test or the "teaching-suggestion-motivation" test—that further lowered the effective standard of invention.[116] In obviousness cases, the defendant typically argues that the patented invention is a *combination* of elements that are found separately in the prior art, and a key question is accordingly whether combining these elements to form the invention would have been an obvious idea. The "suggestion" test limits the evidence that can be used to answer this question. In particular, it emphasizes the need for written materials—like patents or published technical articles—dated before the discovery of the invention, that "suggest[]" (or "teach[]" or "motivate[]") the combination.[117] And by the same token, it limits arguments based on non-written evidence, such as testimony about the general knowledge and common sense of technical workers in the field or the commercial incentives that existed at the time of the invention.

Like many rules of evidence, the "suggestion" test comes with a tradeoff.[118] On the one hand, it reduces the danger of "hindsight bias," that is, the chance that a judge or jury will incorrectly conclude that the invention was obvious.[119] On the other hand, it demands evidence that may not exist in some cases. This is

113 South Corp. v. United States, 690 F.2d 1368 (Fed. Cir. 1982) (en banc).

114 *See* Keating, *supra* note 97, at 262–63 ("*A & P* and *Cuno* are not even suggested by this court.").

115 *See, e.g.*, Union Carbide Corp. v. Am. Can Co., 724 F.2d 1567, 1574 n.4 (Fed. Cir. 1984) (referring to Supreme Court patent doctrines as "cliché[s] which this court has rejected").

116 Duffy, *supra* note 16, at 62–65; *see also* Robert W. Harris, *Prospects for Supreme Court Review of the Federal Circuit Standards for Obviousness of Inventions Combining Old Elements*, 68 J. PAT. & TRADEMARK OFF. SOC'Y 66, 71–74 (1986) [hereinafter Harris, *Prospects for Supreme Court Review*]; Robert W. Harris, *Critique of the Federal Circuit's Suggestion Test for Obviousness*, 72 J. PAT. & TRADEMARK OFF. SOC'Y 990, 991 (1990) [hereinafter Harris, *Critique of the Federal Circuit's Suggestion Test*].

117 KSR Int'l Co. v. Teleflex Inc., 550 U.S. 398, 419 (2007) (noting that the test emphasized "published articles and the explicit content of issued patents").

118 *See* Duffy, *supra* note 16, at 63–65.

119 *See id.* at 64–65.

particularly likely to be true in fast-moving areas of new technology, where obvious ideas "in the air" have not yet been reduced to writing, and inventors instead exploit new opportunities by following common sense and commercial incentives.

Much like earlier changes in the law, the "suggestion" test was established incrementally through a series of cases. It was first mentioned in Federal Circuit opinions issued in 1983.[120] The test was then applied and developed in a series of cases throughout the 1980s, and by the end of the decade, commenters were prepared to recognize that it had crystalized into a new rule of "general applicability" in patent cases that represented a "more liberal, pro-patentee" approach.[121] The "suggestion" test remained the governing rule through the early 2000s (although some commenters have argued that different Federal Circuit opinions seemed to take different positions on just how strictly the test should be applied).[122]

F. *KSR International Co. v. Teleflex Inc.*

The sixth (and final) era of invention doctrine was inaugurated by renewed fears and complaints in the late 1990s and early 2000s that strong patent rights once again seemed to be hindering, rather than promoting, innovation. Lemley gives the following account:

> These complaints shared a worry that two decades of strengthening patents had led to a wave of bad patents approved by the PTO in the 1990s and asserted in the 2000s; to a pervasive problem of patent holdup as companies faced hundreds of suits, each with the potential to shut down its core product; and to a flood of suits by nonpracticing entities (NPEs), or "patent trolls," that took advantage of plaintiff-friendly rules to extract more money than they deserved.[123]

---

120 *See* Harris, *Prospects for Supreme Court Review*, *supra* note 116, at 72.

121 Harris, *Critique of the Federal Circuit's Suggestion Test*, *supra* note 116; *see also* Harmon, *supra* note 107, at 575 ("[T]he task facing the patent challenger has been made even more difficult by the court's insistence, in a dozen or more cases, that there must be something in the prior art as a whole to suggest the desirability, and thus the obviousness, of making a combination of prior art references. It is impossible for a patent lawyer of reasonable experience to read these cases without concluding that obviousness is a much more problematic defense than it was a decade ago." (footnote omitted)); Dreyfuss, *supra* note 110, at 16 (commenting, in 1989, that the Federal Circuit had begun to apply a new "teaching, suggestion, or inference" test).

122 *See, e.g.*, Joshua McGuire, *Nonobviousness: Limitations on Evidentiary Support*, 18 BERKELEY TECH. L.J. 175, 179 (2003); Rebecca S. Eisenberg, *Obvious to Whom? Evaluating Inventions from the Perspective of PHOSITA*, 19 BERKELEY TECH. L.J. 885, 894–96 (2004).

123 Lemley, *supra* note 1, at 9–10 (footnotes omitted).

These concerns were captured in a pair of reports—one from the Federal Trade Commission in 2003,[124] and another from the National Research Council in 2004[125]—both of which called for the Federal Circuit's "suggestion" test to be struck down.

Three years later, the Supreme Court acted. Its 2007 opinion, *KSR International Co. v. Teleflex Inc.*, reversed the "suggestion" test, reasoning that "[t]he obviousness analysis cannot be confined by a formalistic conception of the words teaching, suggestion, and motivation, or by overemphasis on the importance of published articles and the explicit content of issued patents," and that "[g]ranting patent protection to advances that would occur in the ordinary course without real innovation retards progress."[126]

The impact of this decision was largely to unwind the Federal Circuit's case law of the 1980s, and return patent law to its pre-"suggestion" (and pre-"genius") focus on simple "nonobviousness."[127] In the early twenty-first century, patent law asked largely the same question that had been asked by the original late nineteenth- and early twentieth-century "nonobviousness" cases: whether an invention involves "something more than what is obvious to persons skilled in the art to which it relates."[128]

III. TECHNOLOGICAL THEORY

A. Literature

The technological theory is that these shifts in the standard of invention are explained by underlying changes in the "pace of technological advance."[129] This section explains the technology theory, as it has been presented in the existing literature.

The first account of this theory seems to be the one found in Judge William K. Townsend's 1901 history of U.S. patent law in

---

124 FED. TRADE COMM'N, TO PROMOTE INNOVATION: THE PROPER BALANCE OF COMPETITION AND PATENT LAW AND POLICY (2003), https://www.ftc.gov/sites/default/files/documents/reports/promote-innovation-proper-balance-competition-and-patent-law-and-policy/innovationrpt.pdf [https://perma.cc/5VAZ-3DNN].

125 NAT'L RSCH. COUNCIL, A PATENT SYSTEM FOR THE 21ST CENTURY (Stephen A. Merrill et al. eds., 2004).

126 KSR Int'l Co. v. Teleflex Inc., 550 U.S. 398, 419 (2007).

127 Duffy, *supra* note 16, at 68.

128 *Id.* at 41 (quoting Pearce v. Mulford, 102 U.S. 112, 118 (1880)) (noting that the most common late nineteenth- and early twentieth-century formulations were "quite similar to the modern standard," "nearly identical to the modern statute," and "very close to the modern nonobviousness test").

129 Kitch, *supra* note 16, at 304.

the eighteenth and nineteenth centuries.[130] Townsend himself served as a federal district judge in Connecticut during the late nineteenth century (and had experience with patent cases),[131] and his goal was to explain the rise of the new "nonobviousness" regime in the post-Civil War period. Townsend's account of the rise of this doctrine is so interesting (and remains unsurpassed as a firsthand description of these events), that it is worth quoting at length:

> [O]ur courts . . . and the inventor have grown wiser by reason of the lessons of the latter half of the nineteenth century.
>
> Coincident with the development of our industries came the great inventions of the century just closed, steam-power, cotton-gins, looms, manufacturing machinery, india-rubber, mowing-machines, reapers, telegraphs, telephones, electric motors and lights; each primary invention opening up a vast field of undeveloped possibilities. . . .
>
> Then after each principal pioneer invention followed a great host of minor improvements. . . . [I]n this lesser field of minor inventions, artisans, by means of patents for mere modifications, prevented others from making or using obvious improvements on existing devices. . . .
>
> At length, when the inventive field became so crowded with secondary patents that there was no standing room for an inventor, and no pathway open for forward progress to the workman or the public, the courts undertook to find a remedy for this evil. They began to hold that the mere production of something shown not to have been made before did not of itself furnish consideration for a monopoly, but that such production must have required invention in order to entitle it to the protection of the patent law . . . .
>
> Between the years 1880 and 1890, the Supreme Court repeatedly asserted this doctrine, applying it in the case of new forms of manufacture and new combinations of existing devices having superior utility but not requiring inventive conception.[132]

The essence of Townsend's theory is that the problems in the patent system of the late nineteenth century—a "crowded" field of overlapping patents that blocked innovation—were traceable to the new technological opportunities opened up by the "primary invention[s]" of the nineteenth century.[133] In a period of slow progress, like the early nineteenth century, when new ideas were scarce, the patent system could afford to grant patents freely on

---

130 TOWNSEND, *supra* note 16.

131 Indeed, a Townsend opinion is cited by the famous Walker patent law treatise as a leading application of the new nonobviousness doctrine. WALKER, *supra* note 66 ("Judge Townsend has said that an electrical patent should be stripped of the dazzling halo which conventionally adorns appliances designed to deal with that mysterious agent, electricity; when a court is called on to decide the question of the presence or absence of invention, in an electrical patent.").

132 TOWNSEND, *supra* note 16, at 395–96.

133 *Id.*

almost any new device. However, the conditions of technological ferment and rapid technological progress in the late nineteenth century demanded a stricter approach, and the courts reacted.

Certainly, there seemed to be no shortage of complaints about conflicting patent rights in the late nineteenth century. Railroads complained that train cars had become covered by so many overlapping, conflicting patents (on the components of the engines, the brakes, the wheels, and so on) that it was impossible to clear the rights.[134] Farmers complained that the same situation prevailed with the new agricultural machinery—no one could plow a field without the fear of patent infringement, and ruthless "patent sharks" exploited the situation by filing vexatious patent infringement lawsuits against farmers.[135] And, more generally, the new industries of the late nineteenth century (railroads, rubber, electricity, and others) fought a series of protracted "patent wars" in these decades.[136]

At least three later histories of this period seem to have independently come to similar conclusions as Townsend. Kitch's 1966 history of the "nonobviousness" doctrine argues that the adoption of this doctrine in the late nineteenth century could only be explained by the "non-legal forces at work in the country after the Civil War," and pointed to the "quickening pace of technological advance, particularly after the Civil War," reasoning that "as the pace of obsolescence in a society quickens, the standard of invention must be raised lest every common product be the subject of a patent monopoly."[137] B. Zorina Khan's 2004 history of changes in patent institutions during the nineteenth century contends that "[t]he heightened [obviousness] standards likely functioned as a more effective filter in view of the great increase in patenting rates and technical qualifications among the population during the postbellum period."[138] And, lastly, Duffy's

---

134 *E.g.*, Steven W. Usselman, *Patents Purloined: Railroads, Inventors, and the Diffusion of Innovation in 19th-Century America*, 32 TECH. & CULTURE 1047, 1063 (1991) ("[T]he Patent Office had created a complex web of claims that would baffle attorneys for decades."); STEVEN W. USSELMAN, REGULATING RAILROAD INNOVATION: BUSINESS, TECHNOLOGY, AND POLITICS IN AMERICA, 1840–1920, at 104–10 (2002); Christopher Beauchamp, *The First Patent Litigation Explosion*, 125 YALE L.J. 848, 929–30 (2016).

135 *E.g.*, Earl W. Hayter, *The Patent System and Agrarian Discontent, 1875–1888*, 34 MISS. VALLEY HIST. REV. 59, 62–66 (1947) ("The farmers had neither the time, money, nor skill 'to wade through this vast labyrinth of inventions.'"); Beauchamp, *supra* note 134, at 925–26.

136 B. Zorina Khan, *Trolls and Other Patent Inventions: Economic History and the Patent Controversy in the Twenty-First Century*, 21 GEO. MASON L. REV. 825, 841 (2014).

137 Kitch, *supra* note 16, at 304–05, 320.

138 Khan, *supra* note 16, at 16–17.

2007 history asserts that "[t]he obviousness doctrine will be least important in societies where: (1) patent rights are expensive to obtain and to enforce, (2) the pace of social change is relatively slow, (3) few inventors are likely to be working on similar projects, and (4) patent rights are kept relatively narrow," and that, as social and technological conditions changed during the nineteenth century, the need for this doctrine grew more dire.[139]

These accounts are focused on the late nineteenth century, but other evidence suggests that the same dynamics—major technological breakthroughs followed by periods of concern about conflicting rights in the patent system—have occurred in the twentieth and twenty-first centuries as well. For example, economic historians David Mowery and Nathan Rosenberg argue that technological progress during the middle decades of the twentieth century was driven in substantial part by three "central clusters" of basic innovations: "the internal combustion engine, electricity (including electronics), and chemistry."[140] Contemporary sources reported similar intellectual-property conflicts in these industries, and drew similar lessons. For example, Sylvester Petro, after describing intractable patent conflicts in the new chemicals industry,[141] concluded: "In a developing technology there may be hundreds of important patents to be reckoned with," and "[i]t is often impossible to tell which patents are the key patents in a complex and rapidly growing art. . . ."[142] The result is that "the holder of any patent, if he has weight and swings it properly, may block a whole line of development; and even if no conscious attempt at frustration is made, the mere presence of conflicting, legally enforceable rights makes for confusion and obstruction."[143]

---

139 Duffy, *supra* note 16, at 17–18.

140 David Mowery & Nathan Rosenberg, *Twentieth-Century Technological Change*, *in* 3 THE CAMBRIDGE ECONOMIC HISTORY OF THE UNITED STATES: THE TWENTIETH CENTURY 803, 807 (Stanley L. Engerman & Robert E. Gallman eds., 2000); *see also* Alexander J. Field, *The Most Technologically Progressive Decade of the Century*, 93 AM. ECON. REV. 1399, 1399 (2003); *Investigation of Concentration of Economic Power: Hearings Before the Temp. Nat'l Econ. Comm.*, 76th Cong. 16207 (1940) (statement of Joseph O'Mahoney, Chairman, Temp. Nat'l Econ. Comm.) ("[T]he 10 years since the crash of 1929 have probably seen the establishment of more new industries and greater technological gains in old industries than any decade since the human race first began to measure time.").

141 Sylvester Petro, *Patents: Judicial Developments and Legislative Proposals*, 12 U. CHI. L. REV. 352, 388 (1945) ("The foregoing situation was simple and innocuous in comparison with the one recently prevailing in the synthetic rubber field.").

142 *Id.* at 390.

143 *Id.*

Similarly, today's revolutionary advances have been concentrated in computers and biotechnology.[144] And modern accounts of problems in the patent system have once again focused on very similar concerns in these industries—"thickets"[145] or "anticommons" of conflicting patent rights that block innovation.[146]

More broadly, historians of technology have often argued that technological progress is driven by major breakthroughs ("macroinventions" or "general purpose technologies") that temporarily open up new opportunities for rapid sequences of cumulative invention.[147] For example, Joel Mokyr argues that "[t]he chief importance of radical inventions is that they raise the marginal product of effort in development, and thus lead to a sequence of further improvements," and "[a]n implication of this theory is that, in periods of radical inventions, we observe an intensification of smaller inventions as well."[148] Eventually, the opportunities created by a basic breakthrough run into diminishing returns, and progress slows, until the next breakthroughs come along to restock the larder. Accordingly, over time, the opportunities for innovation, and the pace of technological progress, tend to fluctuate, driven by the "randomness associated with the discovery of great inventions."[149] The claim of the technological theory is that these underlying movements in technological conditions lie behind the recurring cycles in patent law.

B. Model

The goal of this section is to present a model that captures the essential empirical claim of the technological theory—changes in the standard of invention have consistently followed behind

144 *E.g.*, ALFRED D. CHANDLER JR., INVENTING THE ELECTRONIC CENTURY 132–76 (2001) (describing the "microprocessor revolution"); Klaus Buchholz & John Collins, *The Roots—A Short History of Industrial Microbiology and Biotechnology*, 97 APPLIED MICROBIO. & BIOTECH. 3747, 3757 (2013) ("[A] large number of significant scientific breakthrough events as well as technological progress provided a new basis for [biotechnology].").

145 Carl Shapiro, *Navigating the Patent Thicket: Cross Licenses, Patent Pools, and Standard Setting*, *in* 1 INNOVATION POLICY AND THE ECONOMY 119, 119 (Adam B. Jaffe, Josh Lerner & Scott Stern eds., 2001).

146 Michael A. Heller & Rebecca S. Eisenberg, *Can Patents Deter Innovation? The Anticommons in Biomedical Research*, 280 SCI. 698, 698 (1998).

147 *E.g.*, Mokyr, *supra* note 17, at 351–52; Bresnahan & Trajtenberg, *supra* note 17, at 84; *see also, e.g.*, Zvi Griliches**,** William D. Nordhaus & F. M. Scherer, *Patents: Recent Trends and Puzzles*, 1989 BROOKINGS PAPERS ECON. ACTIVITY: MICROECON. 291, 324–25 (Nordhaus comments); Nathan Rosenberg, *Technological Interdependence in the American Economy*, 20 TECH. & CULTURE 25, 26–32 (1979).

148 Mokyr, *supra* note 17, at 352.

149 *See* Griliches, *supra* note 147.

underlying changes in the pace of technological progress. There are two key features of the legal history, as described in the prior section,[150] that any adequate model of this claim would need to capture:

1. **The Role of Precedent**. The theory is that the standard of invention has been influenced by the rate of technological progress. However, changes in the rate of technological progress have not *instantaneously* translated into legal change. Instead, the legal standard has evolved gradually, through a common-law-like legal process, where judges are influenced by the existing body of precedent.[151] In other words, in each year, judges give some weight to established precedents and only gradually and incrementally depart from them as new conditions require. For example, Townsend tells us that, in the late nineteenth century, the standard of invention did eventually rise under the pressure of new technological conditions, but only "[a]t length."[152] Accordingly, the legal standard applied in a particular year is generally not independent of the law in the prior year. Instead, the legal standard applied by judges in any given year will generally be a *mixture* of both: (a) the standard established by prior cases (which carry the force of precedent), and (b) current technological conditions (which may push the law in new directions).
2. **The Available Observations of Legal Change**. Although the law has evolved through a series of small, incremental steps, which gradually raise or lower the legal standard, a challenge is that it is not generally possible to precisely quantify the incremental change in the law in each year, i.e., to say that the standard of invention in 1872 was precisely X increments higher than the standard in 1871.[153] Instead, the available "observations" of legal change are the broader judgments made by doctrinal commenters about when the process of gradual legal change has gone far enough to qualify as a change in regime. In other words, these commenters watch the development of the law and report when the doctrine has departed sufficiently far from the old baseline to enter a new regime; for example, in the late nineteenth century,

150 *See supra* Part II.
151 *See supra* Part II.
152 *See* Townsend, *supra* note 16, at 396.
153 *See supra* Part II.

> the law transitioned from the "substantial novelty" regime to the "nonobviousness" regime. And the property these reported regimes have is a *relative* ranking as stricter or more lenient.

A natural framework, capable of capturing both of these features, is the dynamic ordered probit.[154] This framework was originally introduced by Barry Eichengreen, Mark Watson, and Richard Grossman to study the influence of economic conditions on central-bank policy decisions and has been extended by a number of other papers.[155] It has never been used to model the evolution of a legal standard. However, the key features of the problems are relevantly similar. In particular, as applied here, the model has three essential parts.

The first is the actual standard of invention applied by courts in year **t**, which can vary continuously over time. In other words, if a patent lawsuit were filed, and a judge (or jury) had to decide whether the asserted patent was valid, this is the standard that would be applied. This standard (y_t^*) is defined as follows:

$$y_t^* = \theta y_{t-1}^* + \beta R_t + \varepsilon_t, \quad \varepsilon_t \sim N(0,1), i.i.d.$$

The definition says that the legal standard applied in each year is a function of three things: (1) the legal standard applied in the prior year (y_{t^*-1}), whose precedential force is determined by the coefficient θ (theta), (2) the current rate of technological progress (R_t), whose influence on the law is determined by the coefficient β (beta), and (3) a random error (ε_t) (epsilon). This definition captures the process of common-law legal change—it allows movements in the rate of technological progress to gradually influence the direction in which the doctrine evolves over time. As a stationarity condition (and realistic description of the force of precedent), $\theta < 1$.

The second part of the model is the legal regime reported by doctrinal commenters in year **t** (Y_t). Commenters have used three core legal regimes to describe the evolution of the standard of invention: (1) "substantial novelty" (lowest standard); (2) "nonobviousness" (intermediate standard); and (3) "genius"

---

154 *See* Eichengreen, Watson & Grossman, *supra* note 30, at 733–35.

155 *See id.*; *see also* Michael Dueker, *Conditional Heteroscedasticity in Qualitative Response Models of Time Series: A Gibbs-Sampling Approach to the Bank Prime Rate*, 17 J. BUS. & ECON. STAT. 466, 466–69 (1999); Marcelle Chauvet & Simon Potter, *Forecasting Recessions Using the Yield Curve*, 24 J. FORECASTING 77, 77–85 (2005); Sebastian Fossati, *Dating U.S. Business Cycles with Macro Factors*, 20 STUD. NONLINEAR DYNAMICS & ECONOMETRICS 529, 534 (2016).

(highest standard).[156] However, as the earlier historical discussion also illustrated, these three categories are not quite a rich enough language to capture the judgments made by doctrinal commenters.[157] In particular, commenters have often described the law as spending extended periods of time *in between* these regimes—for example, the period of gradual transition from "substantial novelty" to "nonobviousness" in the late nineteenth century, or the period of division between the circuits following the passage of the Patent Act of 1952. Accordingly, the model will use five categories to represent the legal judgments made by doctrinal commenters about the state of the law in each year (labeled 1–5), which allow the doctrine to be in one of three core legal regimes or between them:

1 = "substantial novelty"

2 = in between "substantial novelty" and "nonobviousness"

3 = "nonobviousness"

4 = in between "nonobviousness" and "genius"

5 = "genius"

Finally, there is a set of four cutpoints (c_1, c_2, c_3, c_4) that map the exact legal standard applied by courts in each year (y_t^*) to the legal regime reported by doctrinal commenters (Y_t):

$Y_t = 1$ if $y_t^* < c_1$

$Y_t = 2$ if $c_1 \leq y_t^* < c_2$

$Y_t = 3$ if $c_2 \leq y_t^* < c_3$

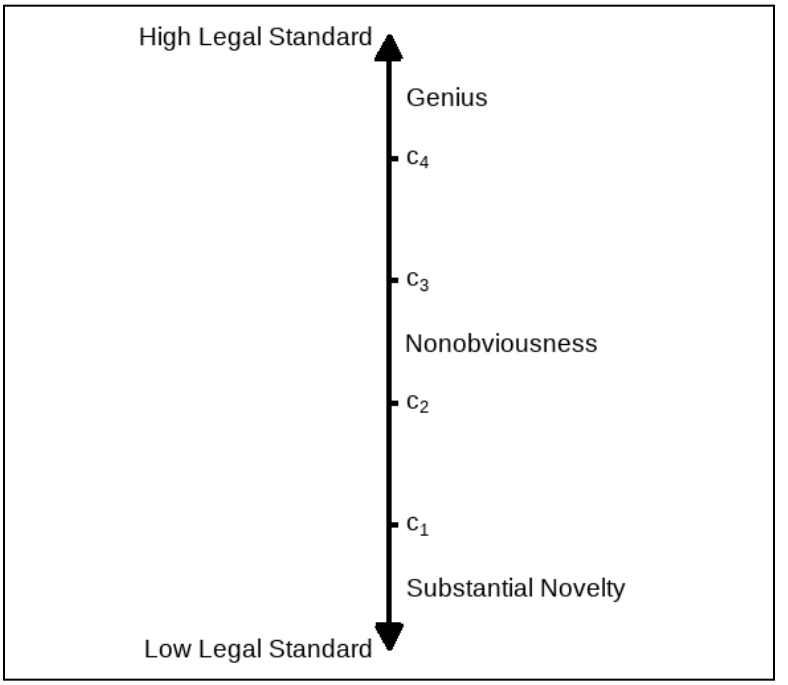
$Y_t = 4$ if $c_3 \leq y_t^* < c_4$

$Y_t = 5$ if $c_4 \leq y_t^*$

Figure 1 (below) represents these cutpoints visually. The cutpoints capture the nature of the observation process. Doctrinal commenters observe the evolution of the legal standard over time and classify the law in each year as falling into a particular legal regime based on the strictness of the standard applied by courts.[158] In so doing, commenters distinguish between "minor" fluctuations in the legal standard, which are not large enough to qualify as a regime change, and "major" movements, which do qualify as a jump into a new regime.[159] The cutpoints define *how much* movement in the underlying legal standard (y_t^*) is enough to qualify as a change in regime (Y_t).

156 *See supra* Part II.
157 *See supra* Part II.
158 *See supra* Part II.
159 *See, e.g.*, Prager, *supra* note 64, at 81.

Figure 1: Cutpoints



Within this modeling framework, the data are: (1) the rate of technological progress in each year (R_t), and (2) the legal regime reported by doctrinal commenters in each year (Y_t). And the parameters to estimate are the coefficient on the rate of technological progress (β), the coefficient on the prior legal standard (θ), and the cutpoints (c_1, c_2, c_3, c_4).

V. DATA

A. Rate of Technological Progress

To measure the rate of technological progress in each year (R_t), this Article relies on the two "traditional,"[160] "standard"[161] long-run, quantitative measures of technological progress that have been developed by economic historians: (1) rate of productivity growth, and (2) number of technologically significant patents issued by the USPTO.

---

160 *See* ROBERT J. GORDON, THE RISE AND FALL OF AMERICAN GROWTH 537 (2016) (noting that productivity growth is the "traditional" general-purpose measure of the pace of technological progress).

161 *See* Moser, *supra* note 2, at 23–24 ("[P]atent counts have become the standard measure of innovation.").

The rate of productivity growth measures changes in the efficiency of the economy over time, which is a measure of better technology in the broad sense.[162] And each invention issued by the USPTO represents, at least in theory, a new technology. Of course, not all patents are created equal—some are important, most are irrelevant. The standard measure of technological significance among patents is citation count (more highly cited patents are likely to be more technologically central).[163] A challenge is that citation data is only available for recent decades. However, a new line of work measures technological significance in a new way, using high-dimensional text analysis to identify patents that introduce new, influential, technical terms.[164] For example, by this metric, the first patent to introduce the term "alternating current" would score highly—the term is both novel (not used in prior patents) and influential (used in many future patents)—which captures basic new ideas entering the patent system. This measure (dubbed "breakthrough patents")[165] is available over the long term, reaching back to the mid-nineteenth century. And the number of these patents in force in each year plausibly captures the mechanism of the technological theory—clusters of "primary invention[s]" that create new technological opportunities.[166]

In particular, the two measures of technological progress the Article will use are:

1. The rate of productivity growth in each year
2. The number of "breakthrough patents" in force in each year (scaled per capita)

The merits of productivity growth and patenting as measures of technological progress have been much debated, and neither measure is perfect.[167] For example, productivity growth likely captures a somewhat broader notion of technology than the universe of merely *patentable* technology. And text-based measures of technological significance need to account for complex changes in the way language has been used across a two-century patent corpus. However, reassuringly, the two measures are

---

162 *See* Nicholas Crafts & Pieter Woltjer, *Growth Accounting in Economic History: Findings, Lessons and New Directions*, 35 J. ECON. SURVS. 670, 672 (2021) ("[Total Factor Productivity] is usually interpreted as a measure of technology, summarizing how intensively and efficiently inputs are used in production.").

163 *See* Moser, *supra* note 2, at 23–24.

164 Bryan Kelly et al., *Measuring Technological Innovation over the Long Run*, 3 AM. ECON. REV.: INSIGHTS 303, 303 (2021).

165 *Id.* at 305.

166 Townsend, *supra* note 16, at 395; *see supra* Section III.A.

167 *See generally* Crafts & Woltjer, *supra* note 162, at 670–71; Moser, *supra* note 2, at 23–24.

positively correlated—a higher number of technologically significant patents predicts higher productivity growth.[168] To the extent these measures agree, it is plausible that they are picking up a shared notion of technological progress that is both economically meaningful and relevant to the patent system.

The full details of the data are described in Appendix A. Productivity data for the nineteenth century is available from Abramovitz,[169] and productivity data for the twentieth century is available from three sources: (1) Alexander J. Field,[170] (2) Robert J. Gordon,[171] and (3) Gerben Bakker, Nicholas Crafts, and Pieter Woltjer.[172] Accordingly, there are three potential productivity series available for the entire period, using the Abramovitz data for the nineteenth century and one of three possible sources for the twentieth century:

1. Abramovitz + Field
2. Abramovitz + Gordon
3. Abramovitz + Bakker

In total, each series covers the period 1800–2007. For the reasons described in Appendix A, the Field data is the best fit for the Article's analysis. However, the Article will estimate model with the other series as robustness checks. Data for "breakthrough patents" is available from Kelly et al.[173] The series for the number of "breakthrough patents" in force runs from 1856–2002.

B. Doctrinal Coding

To measure the legal regime in each year (Y_t), the Article codes the law in each year as falling into one of the five categories of the model:

1 = "substantial novelty"

2 = between "substantial novelty" and "nonobviousness"

3 = "nonobviousness"

4 = between "nonobviousness" and "genius"

5 = "genius"

---

168 *See* Kelly et al., *supra* note 164.

169 Moses Abramovitz, *The Search for the Sources of Growth: Areas of Ignorance, Old and New*, 53 J. ECON. HIST. 217, 228 tbl.2 (1993).

170 ALEXANDER J. FIELD, A GREAT LEAP FORWARD 149 tbl.6.1 (2011).

171 Robert J. Gordon, *Revisiting U.S. Productivity Growth over the Past Century with a View of the Future* 44 tbl.8 (Nat'l Bureau of Econ. Rsch., Working Paper No. 15834, 2010).

172 Gerben Bakker, Nicholas Crafts & Pieter Woltjer, *The Sources of Growth in a Technologically Progressive Economy: The United States, 1899–1941*, 129 ECON. J. 2267, 2280 tbl.4 (2019).

173 Kelly et al., *supra* note 164, at 315 fig.4.

To carry out the coding, the Article relies on the judgments of doctrinal commenters about the state of the law in each year, which were described in the history section above.[174] Based on that history, most coding decisions are straightforward. However, three questions deserve special comment. This section will discuss each one, then present the coding scheme as a whole.

1. Transition Dates

One challenge is that many of the transitions between legal regimes were gradual, incremental processes accomplished through a series of cases.[175] Accordingly, there is sometimes room for reasonable disagreement about the precise beginning or ending dates of these transitions. In particular, there are four such transition dates where the record of doctrinal commentary leaves open a range of possible views:

- The beginning of the transition to "nonobviousness" in the late nineteenth century might have occurred as early as 1870 or as late as 1875.[176]
- The end of the transition to "nonobviousness" in the late nineteenth century might have occurred as early as 1880 or as late as 1885.[177]
- The beginning of the transition to "genius" in the early twentieth century is most commonly described as occurring around 1930 but has been dated as early as 1925 or as late as 1939.[178]
- The Federal Circuit's "suggestion" test first appeared in cases in 1983, was then gradually developed through a series of cases, and had clearly ripened into a new "general" rule by the end of the 1980s—and therefore might be regarded as established at any point in between 1983 and 1990.[179]

This Article's approach to these timing questions is the following. First, it creates a "main" coding scheme, using the central values of the ranges identified by doctrinal commentary. For example, the beginning of the nineteenth-century transition to "nonobviousness" is coded as occurring in 1873 (between 1870

174 *See supra* Part II.
175 *See supra* Part II.
176 *See supra* Section II.B.
177 *See supra* Section II.B.
178 *See supra* Section II.C.
179 *See supra* Section II.E.

and 1875). Second, the Article then creates pairs of "alternative" coding choices for each date, capturing the outer boundaries of uncertainty. For example, the beginning of the nineteenth century transition to "nonobviousness" is coded as occurring in either 1870 or 1875—the earliest and latest dates identified by commentary.

2. Post-1952 Circuit Divisions

A second question is how to code the decades following the enactment of the Patent Act of 1952. The key feature of this period was persistent division between the circuit courts of appeal.[180] Some circuits subscribed to the view that the 1952 Act had overturned the strict Supreme Court precedents of the prior decades and wound the clock back to the pre-1930 era (Category 3 in Table 1, located at the end of this section), while other circuits held that this legislation had instead largely codified the strict legal standard of the 1940s (Category 5). And litigants might face different legal rules depending on the circuit in which their case was filed.

This Article's approach is to code these years as Category 4, placing them somewhere between Category 3 and Category 5. This coding choice captures the "average" or "expected" state of the law. In these years, litigants competed to get cases filed in their favored circuits but had some chance of encountering either regime.

3. The Federal Circuit's "Suggestion" Test

The final question is where to place the Federal Circuit's "suggestion" test in the coding scheme. The source of uncertainty here is simply thin commentary. This test is generally described as a lenient, "pro-patentee" version of "nonobviousness."[181] However, to the author's knowledge, no commentary tries to comprehensively rank and compare it to the other standards. For example, Duffy's commentary contains a table setting forth and ranking the core legal regimes used by courts, but the "suggestion" test does not have its own entry in this table.[182]

This Article's approach is the following. First, by far the most natural coding choice is to place the "suggestion" test in Category 2, i.e., somewhere between "substantial novelty" (Category 1) and "nonobviousness" (Category 3). This coding choice captures the idea that the "suggestion" test is a more lenient, pro-patentee

180 *See supra* Section II.C.

181 *E.g.*, Harris, *Critique of the Federal Circuit's Suggestion Test*, *supra* note 116.

182 Duffy, *supra* note 16, at 18.

version of the "nonobviousness" regime, but does not go so far as to overturn the "nonobviousness" regime altogether. The Article uses this coding choice for the main coding scheme.

However, as alternatives, it also considers Category 1 and Category 3. At one end of the spectrum, perhaps it could be argued that the "suggestion" test was so extreme, and limited the available evidence of "nonobviousness" so extensively, as to be effectively indistinguishable from the "substantial novelty" regime of the early nineteenth century (and is therefore properly coded as Category 1). Or, at the other end of the spectrum, perhaps it could be argued that the "suggestion" test had little practical importance—it was a mere rule of evidence—and was therefore hardly distinguishable from normal "nonobviousness" (and is properly coded as Category 3). To be clear, in this author's view, neither choice seems like a plausible judgment about the doctrinal history. The "suggestion" test is stricter than normal "nonobviousness," but still maintains the core requirement for originality (unlike "substantial novelty"). However, the alternate coding choice makes it possible to check what the implications of other judgments would be.

4. Coding Scheme

The main coding scheme is shown in Table 1 below. It covers the years 1800–2007, since that is the time period for which technological data is available.[183]

Additionally, there are five total pairs of potential alternative coding choices[184]:

- The beginning of the nineteenth century transition to "nonobviousness" could be coded as occurring in 1870 or 1875.
- The end of the nineteenth century transition to "nonobviousness" could be coded as occurring in either 1880 or 1885.
- The beginning of the twentieth century transition to "genius" could be coded as occurring in either 1925 or 1939.
- The Federal Circuit's "suggestion" test could be coded as established in either 1983 or 1990.
- The Federal Circuit's "suggestion" test could be coded as either Category 1 or Category 3.

183 *See supra* Section IV.A.
184 *See supra* Section IV.B.1–3.

One wrinkle is that the results could be insensitive to any *individual* alternative coding choice but would be sensitive to *combinations* of them. Accordingly, this Article checks the combinations. The set of combinations is listed in Appendix A.

Table 1: Main Coding Scheme

Time Period	Category	Comments
1800 – 1872	1	"Substantial Novelty" Era
1873 – 1882	2	Transition to "Nonobviousness"
1883 – 1929	3	"Nonobviousness" Era
1930 – 1940	4	Transition to "Genius"
1941 – 1951	5	"Genius" Era
1952 – 1981	4	Post-1952 Division Between the Circuits
1982 – 1986	3	Transition to the "Suggestion" Test
1987 – 2006	2	Federal Circuit's "Suggestion" Era
2007	3	Return to "Nonobviousness" with *KSR*

V. ANALYSIS

This section reports the results. Estimation of the model is carried out using the Bayesian, Gibbs-sampler approach for a dynamic probit.[185] The details can be found in Appendix B. The analysis below reports the central estimate and the ninety-five percent confidence interval for each parameter of interest.

In particular, Section V.A presents results where the rate of productivity growth is used as the measure of the rate of technological progress (R_t), and Section V.B

185 *See* Dueker, *supra* note 155, at 471–72; Chauvet & Potter, *supra* note 155 at 85–87, 101–02; Fossati, *supra* note 155, app. A.

presents results where the number of technologically significant patents in force is taken as the measure of the rate of technological progress.

A. Productivity Growth

Table 2 presents results where the model is estimated using the two main series: (a) the Field + Abramovitz productivity series to measure R_t, and (b) the main coding scheme shown in Table 1 to measure Y_t. The estimates show that the legal standard has been positively correlated both with its own past state (central coefficient estimate = 0.957) and with the current rate of technological progress (central coefficient estimate = 0.590).

Table 2: Abramovitz + Field

Parameter	Central Estimate	95% Confidence Interval
ß	0.590	0.262 – 0.940
θ	0.957	0.933 – 0.982
c_1	6.0	1.6 – 11.5
c_2	13.3	7.7 – 19.8
c_3	20.3	13.7 – 28.5
c_4	29.3	20.7 – 39.5

As discussed above, the model is intended to capture a process of gradual legal change.[186] When the long-term rate of technological progress rises, it slowly pushes the legal standard upward. Eventually, over time, the standard reaches a new equilibrium, and the difference between the old and new standards may (or may not) be large enough to qualify as a change in legal regime.

The coefficient estimates in Table 2 determine these equilibria, which represent the long-run effect of a change in the

186 *See supra* Section III.B.

rate of technological progress. In particular, each equilibrium is a pair of points—a prevailing legal standard ($y_t^*{}_{-1}$) and a rate of technological progress (R_t)—such that the expected change in the legal standard is zero, i.e.:

$$E(y_t^*) = \theta y_{t-1}^* + \beta R_t = y_{t-1}^*$$

A little algebra shows that these equilibrium pairs are:

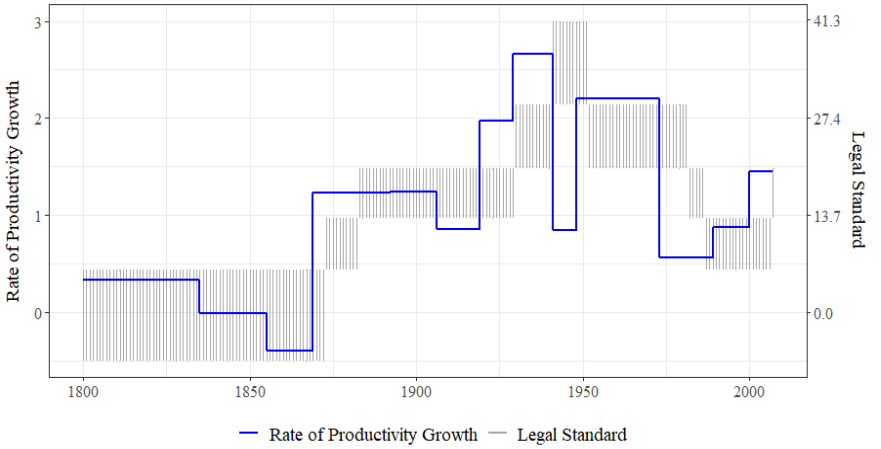
$$y_{t-1}^* = \frac{\beta}{1-\theta} R_t = 13.7 R_t$$

In other words, a one-percent rise in the rate of productivity growth is associated with an increase in the legal standard of 13.7. Of course, that value by itself is not meaningful—it is mapped to observable changes in the legal regime by the cutpoint estimates. As Table 2 shows, the average distance between any pair of estimated cutpoints is 7.8, and the maximum distance between any pair of estimated cutpoints is 9. So, over the long run, a one-percent rise in the rate of productivity growth is associated with (at least) one step upward in the legal regime, but often two.

Using the results in Table 2, Figure 2 plots the rate of productivity growth against the legal standard in each year. Specifically, it takes advantage of the estimated cutpoints to represent the legal standard. The cutpoints indicate that, given the observed legal regime, the exact legal standard must at least lie within certain bounds. For example, if the observed legal regime is "nonobviousness" ($Y_t = 3$), then the exact legal standard must be between c_2 (13.3) and c_3 (20.3). Figure 2 plots these bounds for each year. The ratio between the two axes (13.7) is the equilibrium ratio.

As Figure 2 illustrates, the key implication of the estimates in Table 2 is that changes in the standard of invention have broadly tracked the rate of productivity growth throughout substantially the entire two-century-plus history of the U.S. patent system. In each historical era, legal actors have switched to a stricter legal regime after the rate of productivity growth rose, and vice versa, at a lag of roughly a decade.

Figure 2: Abramovitz + Field



These conclusions are largely insensitive to the alternative estimates of productivity growth and the alternative doctrinal coding schemes. To begin, Tables 3 and 4 (on the next page) present results where the model is estimated using the alternative productivity series: (1) Abramovitz + Gordon, and (2) Abramovitz + Bakker. The results are largely unchanged, along all relevant dimensions. In particular, the equilibrium effect of a one-percent change in the rate of technological progress $\left(\frac{\beta}{1-\theta}\right)$ remains similar—slightly less than twice the average cutpoint distance, and larger than the maximum cutpoint distance:

Abramovitz + Gordon: $\frac{\beta}{1-\theta} = 24.8, \quad avg_c = 13.0, \quad max_c = 18.8$

Abramovitz + Bakker: $\frac{\beta}{1-\theta} = 24.8, \quad avg_c = 7.8, \quad max_c = 9.1$

Figures 3 and 4 create plots analogous to Figure 1 for each series, plotting the rate of productivity growth against the legal standard in each year by using the estimates in Tables 3 and 4. In each case, the broad shape of the doctrinal history similarly tracks long-run changes in the rate of productivity growth.

Table 3: Abramovitz + Gordon

Parameter	Central Estimate	95% Confidence Interval
ß	1.64	1.40 – 2.38
θ	0.934	0.914 – 0.953
c_1	7.3	3.5 – 11.7
c_2	18.8	12.0 – 27.3
c_3	27.4	18.9 – 38.0
c_4	46.3	31.6 – 64.8

Table 4: Abramovitz + Bakker

Parameter	Central Estimate	95% Confidence Interval
ß	0.868	0.401 – 1.38
θ	0.940	0.908 – 0.972
c_1	5.4	1.7 – 9.9
c_2	12.5	7.4 – 18.4
c_3	19.6	13.4 – 26.7
c_4	28.7	20.5 – 38.2

Figure 3: Abramovitz + Gordon

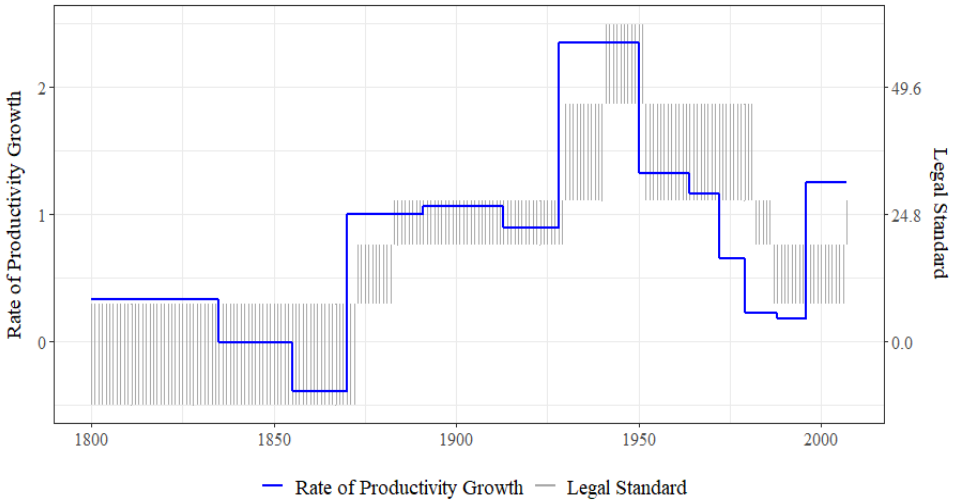
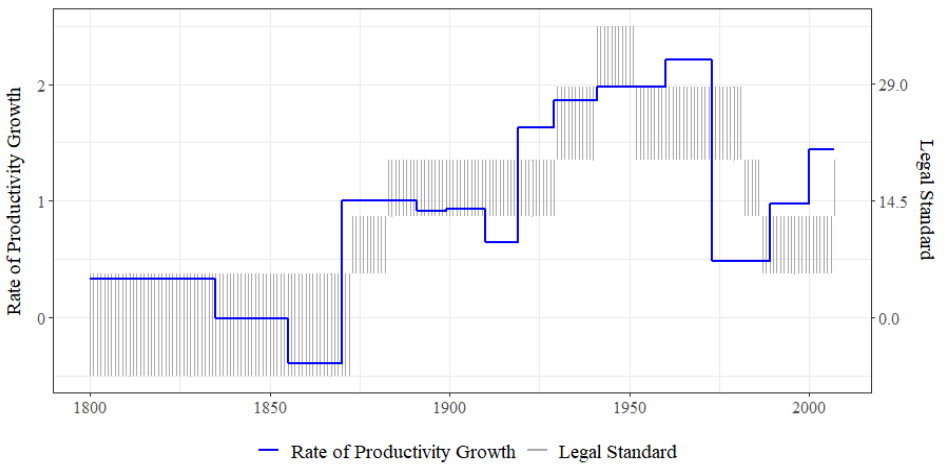


Figure 4: Abramovitz + Bakker



Finally, Table 7 (available in Appendix C, Supplemental Tables) presents results where the model is estimated using the alternative doctrinal-history coding schemes and the main Abramovitz + Field series to measure R_t. As the Table shows, the parameter estimates are largely insensitive to the (small) margins of disagreement among doctrinal commenters.

B. Technologically Significant Patents

Table 5 presents results where the model is estimated using: (a) the number of "breakthrough patents" in force (per 5,000 people) to measure R_t, and (b) the main coding scheme presented in Table 1 to measure Y_t. Again, the estimates show that the legal standard has been positively correlated both with its own past state (central coefficient estimate = 0.980) and with the current rate of technological progress (central coefficient estimate = 0.180).

Table 5: Breakthrough Patents

Parameter	Central Estimate	95% Confidence Interval
ß	0.180	0.031 – 0.344
θ	0.980	0.959 – 0.997
c_1	4.8	0.0 – 10.9
c_2	11.2	5.3 – 18.4
c_3	17.2	10.8 – 24.7
c_4	24.6	17.1 – 33.1

The long-run equilibrium ratio is given below:

$$y^*_{t-1} = \frac{\beta}{1-\theta} R_t = 9\, R_t$$

It shows that one additional breakthrough patent in force per 5,000 people has been associated, over the long-run, with an increase in the legal standard of 9. As Table 5 shows, the average distance between any pair of estimated cutpoints is 6.6, and the maximum distance between any pair of estimated cutpoints is 7.4. Accordingly, over the long run, one additional breakthrough patent in force per 5,000 people has been associated with at least one step upward in the legal regime, and sometimes two.

Figure 5 creates the analogous plot for this series, using the estimates in Table 5. As the Figure illustrates, there are two key points. First, the two measures of technological change—productivity growth and technologically significant patents—closely track each

other (and the doctrine) for most of the history, from the middle of the nineteenth century to the middle of the twentieth century. Second, these two measures then diverge somewhat in the closing decades of the twentieth century (1980s to 2000s). In these decades, the two measures disagree about the size of the most recent wave of technological progress—the productivity data shows a smaller wave, while the patenting data shows a much larger wave. The productivity data is a better match for the doctrine in these decades.

Figure 5: Technologically Significant Patents

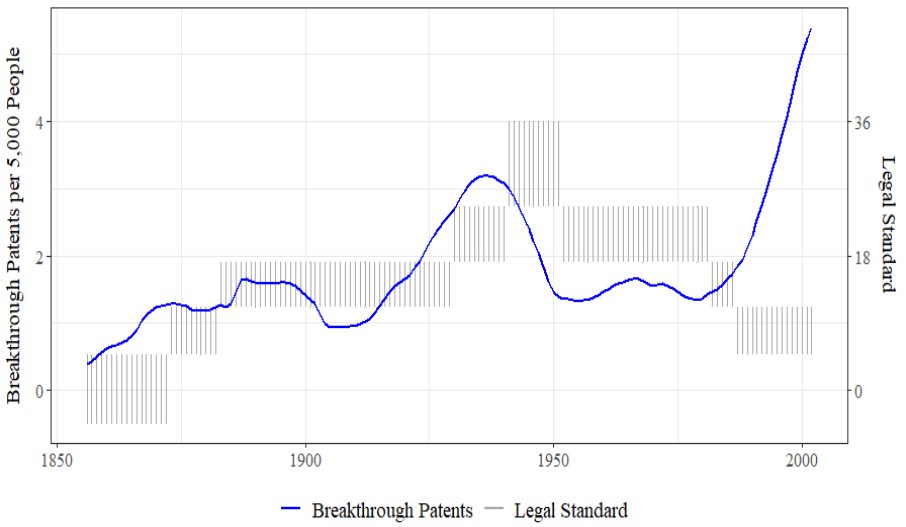


Table 8 (available in Appendix C, Supplemental Tables) presents results where the model is estimated using the alternative doctrinal-history coding schemes. Again, there are two key points. First, the results are largely insensitive to the robustness checks regarding the *timing* of the transitions between regimes (recodings 1–16). Second, relative to the productivity data, the results are much more sensitive to the robustness checks regarding the coding of the Federal Circuit's "suggestion" test (which was the governing test from the 1980s to the early 2000s). In the main coding scheme (Table 1), the "suggestion" test is coded as category 2 (in between "substantial novelty" and "nonobviousness"). However, (a) if the "suggestion" test is coded as category 1 (lower legal standard, recodings 17–32), then the estimated coefficient on the rate of technological progress falls by

roughly a factor of two (average value = 0.0992), and the ninety-five percent confidence interval includes zero; but (b) if the "suggestion" test is coded as category 3 (higher legal standard, recodings 33–40), then the coefficient on the rate of technological progress increases by roughly a factor of two (average value = 0.338). In short, the patenting data is a substantially better fit for the doctrinal history if the history is coded as involving a higher legal standard during the 1980s, 1990s, and early 2000s (when the patenting data shows a larger wave of technological progress).

At present, it remains somewhat unclear why the two measures of technological progress (productivity growth and breakthrough patents) have recently diverged or what lesson should be drawn from this fact. Part of the reason is simply that textual analysis of patents is a relatively new method. Economic historians have refined estimates of productivity growth over the last six decades.[187] By contrast, there is currently only one published, long-run estimate of the number of technologically significant patents using a text-based method.[188] Recent results suggest that it is possible that the divergence between the two measures will eventually be narrowed or even eliminated.[189] And future work may shed more light on these questions.

V. CONCLUSION

In sum, this Article makes two main contributions. First, it draws a missing link between the history of patent rights and the ways that economic historians have thought about measuring long-term technological progress. It shows that a central legal standard in patent law has tracked these measures of technological progress for nearly two centuries. Second, this connection sheds new light on a longstanding puzzle in patent law. The Article's results provide stronger reason to believe that the infamous cycles in patent law are explained by courts acting as responsible stewards of the patent system, reacting to changes in technology and adopting legal regimes that best encourage technological progress in different social conditions.

---

187 *See* Crafts & Woltjer, *supra* note 162, at 670–91.

188 Kelly et al., *supra* note 164, at 303–18.

189 For example, a recent working paper estimates the number of technologically significant patents using a new, text-based metric ("creative patents") and arrives at conclusions that track the productivity data (and the doctrine) much more closely. *See* Aakash Kalyani, *The Creativity Decline: Evidence from US Patents* 21–23 & fig.6 (Fed. Rsrv. Bank of St. Louis, Working Paper No. 2024-008A, 2024).

APPENDIX A: DATA

This appendix provides details on the three data sources: (1) productivity growth, (2) technologically significant patents, and (3) doctrinal coding.

A. Rate of Productivity Growth

Long-term historical productivity data is available from several sources. By convention, all report average rates of productivity growth between peaks of the business cycle to filter out short-run business cycle fluctuations.[190]

For the early to mid-nineteenth century, data is available from Abramovitz, who reports productivity figures for the period 1800–1890, sliced in two ways:

1. A "long" frame, containing average rates of productivity growth during two long historical intervals: 1800–1855 and 1855–1890.
2. A "short" frame, containing average rates of productivity growth during four shorter historical intervals: 1800–1835, 1835–1855, 1855–1870, and 1870–1890.[191]

This Article will use the "short" frame for two reasons. First, the "long" frame slices the data in a way that is not granular enough for this Article's needs. In particular, it merges the decades dominated by the Civil War (1855–1871) with the post-war decades at the end of the nineteenth century (1871–1890). However, the 1870s contained a critical doctrinal transition (the rise of "nonobviousness").[192] This Article therefore requires a frame that is at least fine-grained enough to separate these two periods (1855–1871 vs. 1871–1890). Second, as Field argues, the "short" frame is also the one that is the most consistent with the values reported by the Kendrick data[193] for the period of history where the two series overlap (the post-Civil War years).[194]

---

190 *See* Bakker, Crafts & Woltjer, *supra* note 172; Alexander Field, The Impact of World War II on the Growth of U.S. Potential Output 13–14 (Apr. 16, 2018) (unpublished working paper), https://www.researchgate.net/profile/Alexander-Field-3/publication/324546656 [https://perma.cc/UX7L-PV2D].

191 Abramovitz, *supra* note 169, at 223 tbl.1, 228 tbl.2.

192 *See supra* Part II.

193 *See generally* JOHN W. KENDRICK, PRODUCTIVITY TRENDS IN THE UNITED STATES (1961) (foundational study).

194 FIELD, *supra* note 170, at 148.

For later years, three recent series are available, compiled by different economic historians (all building on data originally reported by Kendrick, which runs back to 1869):

1. Field[195]: private domestic economy, 1869–2007
2. Gordon[196]: private, non-farm domestic economy, 1891–2007
3. Bakker et al.[197]: private, non-farm domestic economy 1899–2007

This Article will use the Field series as the main estimate for three reasons. First, Field is the series that is the most consistent with Abramovitz. Abramovitz reports productivity figures that include the farm sector, as does Field, while Gordon and Bakker et al. report non-farm figures. Second, including the farm sector is especially desirable for a study of patent rights—farms were not just a historically important area of technological innovation but also a central battleground for disputes over patent rights, in particular.[198] Third, Field argues that the reported values are broadly consistent with the "short" frame reported by Abramovitz.[199] For the period of overlap between the two series (1869–1890), this Article will use the most recent data from Field. Finally, as a check, this Article will also estimate the model using the other two series. Accordingly, in total, there are three historical productivity series:

1. Abramovitz + Field (Abramovitz [1800–1868] / Field [1869–2007])
2. Abramovitz + Gordon (Abramovitz [1800–1890] / Gordon [1891–2007])
3. Abramovitz + Bakker (Abramovitz [1800–1898] / Bakker [1899–2007])[200]

B. Breakthrough Patents

Kelly et al. report the number of "breakthrough patents" issued by the PTO for the period 1840–2002.[201] To obtain the

195 *Id.* at 149 tbl.6.1.

196 Gordon, *supra* note 171.

197 Bakker, Crafts & Woltjer, *supra* note 172.

198 Beauchamp, *supra* note 134, at 925–29.

199 FIELD, *supra* note 170, at 146–65.

200 The Bakker et al. series begins in 1899, while the Abramovitz "short" frame runs to 1890. Abramovitz also reports estimates for the post-1890 period, drawing on the data from Kendrick. This Article uses those values for the years 1891–1898.

201 Kelly et al., *supra* note 164, at 315 fig.4.

number of these patents in force in each year, this Article makes a simple assumption—each patent remained in force for 17 years.

That was the term for most of the relevant history.[202] To be clear, there are two issues that this assumption does not account for. First, in the early decades of the nineteenth century, there was a discretionary element to patent term.[203] Specifically, between 1836 and 1861, the term of a patent was 14 years by default but could be extended (and many were) to 21 years by application to the USPTO on a showing of need. Eventually, in 1861, the term was standardized to an intermediate length (17 years). However, individual patents in the early nineteenth century were in force for either longer or shorter times. Second, a patent owner needs to pay maintenance fees to keep a patent in force. Some patents—although technologically significant—likely lapsed before the end of their terms. However, many of the records from the nineteenth century have been lost.[204] And, more generally, it is infeasible to match the set of an estimated 600,000 breakthrough patents to these administrative outcomes. The Article uses a 17-year term as the best approximation.

Accordingly, the series of patents-in-force runs from 1856–2002. This series is also scaled per capita by the number of breakthrough patents in force per 5,000 people.

C. Doctrinal Coding

The main doctrinal coding scheme is presented in Table 1. Table 6 lists the combinations of alternative coding schemes.

Table 6: Alternative Coding Schemes

Number	Coding Decisions
1	1870 / 1880 / 1925 / 1983 / "Suggestion" = 2
2	1870 / 1880 / 1925 / 1990 / "Suggestion" = 2
3	1870 / 1880 / 1939 / 1983 / "Suggestion" = 2
4	1870 / 1880 / 1939 / 1990 / "Suggestion" = 2
5	1870 / 1885 / 1925 / 1983 / "Suggestion" = 2

202 Simon Lester & Huan Zhu, *Rethinking the Length of Patent Terms*, 34 AM. UNIV. INT'L L. REV. 787, 791–793 (2019).

203 *See id.*

204 *Id.* at 793.

6	1870 / 1885 / 1925 / 1990 / "Suggestion" = 2
7	1870 / 1885 / 1939 / 1983 / "Suggestion" = 2
8	1870 / 1885 / 1939 / 1990 / "Suggestion" = 2
9	1875 / 1880 / 1925 / 1983 / "Suggestion" = 2
10	1875 / 1880 / 1925 / 1990 / "Suggestion" = 2
11	1875 / 1880 / 1939 / 1983 / "Suggestion" = 2
12	1875 / 1880 / 1939 / 1990 / "Suggestion" = 2
13	1875 / 1885 / 1925 / 1983 / "Suggestion" = 2
14	1875 / 1885 / 1925 / 1990 / "Suggestion" = 2
15	1875 / 1885 / 1939 / 1983 / "Suggestion" = 2
16	1875 / 1885 / 1939 / 1990 / "Suggestion" = 2
17	1870 / 1880 / 1925 / 1983 / "Suggestion" = 1
18	1870 / 1880 / 1925 / 1990 / "Suggestion" = 1
19	1870 / 1880 / 1939 / 1983 / "Suggestion" = 1
20	1870 / 1880 / 1939 / 1990 / "Suggestion" = 1
21	1870 / 1885 / 1925 / 1983 / "Suggestion" = 1
22	1870 / 1885 / 1925 / 1990 / "Suggestion" = 1
23	1870 / 1885 / 1939 / 1983 / "Suggestion" = 1
24	1870 / 1885 / 1939 / 1990 / "Suggestion" = 1
25	1875 / 1880 / 1925 / 1983 / "Suggestion" = 1
26	1875 / 1880 / 1925 / 1990 / "Suggestion" = 1
27	1875 / 1880 / 1939 / 1983 / "Suggestion" = 1
28	1875 / 1880 / 1939 / 1990 / "Suggestion" = 1
29	1875 / 1885 / 1925 / 1983 / "Suggestion" = 1
30	1875 / 1885 / 1925 / 1990 / "Suggestion" = 1

31	1875 / 1885 / 1939 / 1983 / "Suggestion" = 1
32	1875 / 1885 / 1939 / 1990 / "Suggestion" = 1
33	1870 / 1880 / 1925 / [N/A] / "Suggestion" = 3
34	1870 / 1880 / 1939 / [N/A] / "Suggestion" = 3
35	1870 / 1885 / 1925 / [N/A] / "Suggestion" = 3
36	1870 / 1885 / 1939 / [N/A] / "Suggestion" = 3
37	1875 / 1880 / 1925 / [N/A] / "Suggestion" = 3
38	1875 / 1880 / 1939 / [N/A] / "Suggestion" = 3
39	1875 / 1885 / 1925 / [N/A] / "Suggestion" = 3
40	1875 / 1885 / 1939 / [N/A] / "Suggestion" = 3

APPENDIX B: ESTIMATION

This appendix describes the estimation strategy. Estimation broadly follows the Bayesian, Gibbs-sampler approach for a dynamic probit.[205] The details for each step are described below.

A. Draws of Latent Values

Call $y_1^{*(i)} \ldots y_T^{*(i)}$ the ith draws of y_t^*, for t = 1 … T.

The initial draws $\left(y_t^{*(1)}\right)$ are taken sequentially from $f\left(y_t^{*(1)} \middle| y_{t-1}^{*(1)}\right) \sim \mathrm{N}\left(\theta y_{t-1}^{*(1)} + \beta R_t, 1\right)$, starting with t = 1, and working upward to t = T. Draws are subject to the truncation that y_t^* lies between the cutpoints for Y_t. For example, if Y_t = 3, then $c_2 \leq y_t^* < c_3$. For simplicity, y_0^* is fixed at 0 (a realistic value in all cases).

Subsequent draws of y_t^* are taken according to the approach described below, following the path set out in Fossati.[206] All draws remain subject to the same truncation, i.e., y_t^* lies between the cutpoints for Y_t.

First, a new draw of the last value (t = T) is taken from $f\left(y_t^{*(i+1)} \middle| y_{t-1}^{*(i)}\right) \sim \mathrm{N}\left(\theta y_{t-1}^{*(i)} + \beta R_t, 1\right)$.

Second, new draws for t = T-1, . . . , 2 are taken using the following property. Define:

$$A_t = E(y_t^*) = \sum_{n=0}^{t-1} \theta^n \beta R_{t-n} = \beta R_t + \theta A_{t-1}$$

$$B_t = Var(y_t^*) = \sum_{n=0}^{t-1} \theta^{2n} = 1 + \theta^2 B_{t-1}$$

205 *See* Dueker, *supra* note 155, at 471–72; *see also* Chauvet & Potter, *supra* note 155, at 85–87, 101–02; Fossati, *supra* note 155, app. A; James H. Albert & Siddhartha Chib, *Bayesian Analysis of Binary and Polychotomous Response Data*, 88 J. AM. STAT. ASS'N 669, 669–73 (1993).

206 *See* Fossati, *supra* note 155, app. A.

Then, $f(y_t^* \mid y_{t+1}^*, y_{t-1}^*) \sim \mathrm{N}\left(\widetilde{\mu_t}, \widetilde{\Sigma_t}\right)$, where:

$$\widetilde{\mu_t} = A_t + \theta \begin{pmatrix} B_t \\ B_{t-1} \end{pmatrix}' \begin{pmatrix} B_{t+1} & \theta^2 B_{t-1} \\ & B_{t-1} \end{pmatrix}^{-1} \begin{pmatrix} y_{t+1}^* - A_{t+1} \\ y_{t-1}^* - A_{t-1} \end{pmatrix}$$

$$\widetilde{\Sigma_t} = B_t - \theta^2 \begin{pmatrix} B_t \\ B_{t-1} \end{pmatrix}' \begin{pmatrix} B_{t+1} & \theta^2 B_{t-1} \\ & B_{t-1} \end{pmatrix}^{-1} \begin{pmatrix} B_t \\ B_{t-1} \end{pmatrix}$$

Using this property, new draws are taken sequentially from $f\left(y_t^{*(i+1)} \middle| y_{t+1}^{*(i+1)}, y_{t-1}^{*(i)}\right)$ starting with T = t-1, and working downward to t = 2.

Finally, a new draw is taken for the first value (t = 1) from $f\left(y_1^{*(i+1)} \middle| y_2^{*(i+1)}\right) \sim \mathrm{N}\left(\widetilde{\mu_1}, \widetilde{\Sigma_1}\right)$, where:

$$\widetilde{\mu_1} = A_1 + \theta B_1 B_2^{-1}(y_2^* - A_2) = A_1 + \frac{\theta}{1+\theta^2}(y_2^* - A_2)$$

$$\widetilde{\Sigma_1} = B_t - \theta^2 B_1 B_2^{-1} B_1 = 1 - \frac{\theta^2}{1+\theta^2}$$

B. Prior and Posterior for β

The initial value of β is set at the maximum likelihood estimate (MLE) for a static ordered probit ($y_t^* = \beta R_t + \varepsilon_t$), and a flat, uninformative prior is adopted for β (given that the MLE is not consistent in this context).[207] The posterior distribution of β is then:

$$\beta \sim N\left(\hat{\beta}, \left(X_\beta' X_\beta\right)^{-1}\right)$$

where:

$$\hat{\beta} = \left(X_\beta' X_\beta\right)^{-1} X_\beta' W_\beta \qquad W_\beta = \begin{pmatrix} y_1^* - \beta R_1 \\ \dots \\ y_T^* - \beta R_T \end{pmatrix} \qquad X_\beta = \begin{pmatrix} y_0^* \\ \dots \\ y_{t-1}^* \end{pmatrix}$$

All draws of θ are subject to the condition that $\theta < 1$.

C. Prior and Posterior for θ

The initial value of θ is set at 0.5, and a flat uninformative prior is adopted for θ (as the initial value does not contain useful

207 *See* Albert & Chib, *supra* note 205; Fossati, *supra* note 155, app. A.

information).[208] The posterior distribution of θ is then:

$$\theta \sim N\left(\widehat{\theta}, (X_\theta' X_\theta)^{-1}\right)$$

where:

$$\hat{\theta} = (X_\theta' X_\theta)^{-1} X_\theta' W_\theta \qquad W_\theta = \begin{pmatrix} y_1^* - \beta R_1 \\ \dots \\ y_T^* - \beta R_T \end{pmatrix} \qquad X_\theta = \begin{pmatrix} y_0^* \\ \dots \\ y_{t-1}^* \end{pmatrix}$$

All draws of θ are subject to the condition that $\theta < 1$.

D. Prior and Posterior for c_1, c_2, c_3, c_4

A flat, uninformative prior is adopted for the cutpoints c_1, c_2, c_3, c_4, and initial values are set at the MLE for a static ordered probit.[209] The posterior distribution for cutpoint c_i is then uniform over the interval:

$$[\max\{\max\{y_t^* : Y_t = i\}, c_{i-1}\}, \min\{\min\{y_*^t : Y_T = i+1\}, c_{i+1}\}]$$

E. Gibbs Sampler

To run the sampler, initial draws of y_t^* were first generated; draws were then taken in the following order: cutpoints, y_t^*, β, θ.

Following Gelman et al.,[210] two diagnostics were calculated for each run: (1) $\hat{R}$ (a measure of convergence, using the split-chain approach of Gelman et al.), and (2) *n_eff* (a measure of effective sample size). Gelman et al. suggest the guidelines $\hat{R} < 1.1$ and *n_eff* > 10 per chain = 100. Here, for all runs, and all parameters, $\hat{R} < 1.05$ and *n_eff* > 150.

For each run, ten chains of 250,000 draws were used by default (with the first 50,000 draws reserved for warmup), and this was typically sufficient to achieve satisfactory convergence and sample-size diagnostics ($\hat{R} < 1.05$ and *n_eff* > 150). However, several of the doctrinal robustness checks required longer chains. In particular, these doctrinal robustness checks (chain-lengths) were:

- Productivity—Recoding 2 (300,000)
- Productivity—Recoding 3 (300,000)
- Productivity—Recoding 7 (350,000)

208 *See* Fossati, *supra* note 155, app. A.

209 Albert & Chib, *supra* note 205, at 673.

210 ANDREW GELMAN ET AL., BAYESIAN DATA ANALYSIS 281–88 (3d ed. 2013) (ebook), http://www.stat.columbia.edu/~gelman/book/BDA3.pdf [https://perma.cc/2V2H-S6D7].

- Productivity—Recoding 19 (300,000)
- Productivity—Recoding 24 (300,000)
- Productivity—Recoding 26 (300,000)
- Productivity—Recoding 32 (350,000)
- Productivity—Recoding 33 (400,000)
- Productivity—Recoding 34 (400,000)
- Productivity—Recoding 35 (400,000)
- Productivity—Recoding 36 (500,000)
- Productivity—Recoding 37 (400,000)
- Productivity—Recoding 38 (600,000)
- Productivity—Recoding 39 (500,000)
- Productivity—Recoding 40 (800,000)
- Patenting—Recoding 33 (400,000)
- Patenting—Recoding 34 (500,000)
- Patenting—Recoding 35 (400,000)
- Patenting—Recoding 36 (500,000)
- Patenting—Recoding 37 (300,000)
- Patenting—Recoding 38 (500,000)
- Patenting—Recoding 39 (300,000)
- Patenting—Recoding 40 (600,000)

Following Gelman et al.,[211] thinning was used to address memory issues created by long chains. In particular, for chains longer than 400,000, only every other draw was stored (and, accordingly, 25,000 draws were reserved for warmup).

For each parameter, the Article reports: (1) the mean of the posterior distribution (the "central estimate"); and (2) a 95% confidence interval, i.e., the 2.5% draw from the posterior distribution and the 97.5% draw from the posterior distribution.[212]

211 *Id.* at 282–83.
212 *Id.* at 267.

APPENDIX C: SUPPLEMENTAL TABLES

This appendix reports estimates for the doctrinal robustness checks. In particular, for each check, it reports an estimate for the coefficient on the rate of technological progress, β. Table 7 presents estimates for productivity growth; Table 8 presents estimates for breakthrough patents.

Table 7: Productivity Growth – Recodings

Recoding	Central Estimate (β)	95% Confidence Interval (β)
1	0.494	0.170 – 0.838
2	0.505	0.174 – 0.859
3	0.334	0.086 – 0.601
4	0.363	0.109 – 0.634
5	0.512	0.181 – 0.864
6	0.513	0.177 – 0.878
7	0.333	0.085 – 0.601
8	0.369	0.111 – 0.644
9	0.583	0.237 – 0.950
10	0.607	0.250 – 0.992
11	0.376	0.118 – 0.650
12	0.410	0.144 – 0.692
13	0.615	0.257 – 0.995
14	0.617	0.259 – 1.01
15	0.381	0.121 – 0.656
16	0.421	0.149 – 0.711
17	0.352	0.0903 – 0.632
18	0.360	0.0974 – 0.643

19	0.271	0.0541 – 0.507
20	0.289	0.0688 – 0.526
21	0.355	0.0970 – 0.633
22	0.355	0.0946 – 0.634
23	0.269	0.0541 – 0.505
24	0.284	0.0647 – 0.520
25	0.454	0.151 – 0.783
26	0.468	0.161 – 0.801
27	0.326	0.0892 – 0.580
28	0.344	0.102 – 0.605
29	0.461	0.162 – 0.789
30	0.456	0.151 – 0.788
31	0.326	0.0902 – 0.582
32	0.344	0.100 – 0.607
33	0.506	0.162 – 0.890
34	0.449	0.159 – 0.766
35	0.497	0.158 – 0.873
36	0.454	0.162 – 0.767
37	0.628	0.241 – 1.06
38	0.513	0.203 – 0.843
39	0.622	0.243 – 1.04
40	0.517	0.208 – 0.850

Table 8: Breakthrough Patents – Recodings

Recoding	Central Estimate (β)	95% Confidence Interval (β)
1	0.185	0.0328 – 0.351
2	0.147	0.00634 – 0.305
3	0.169	0.0248 – 0.327
4	0.142	0.00636 – 0.293
5	0.196	0.0387 – 0.366
6	0.157	0.0114 – 0.320
7	0.178	0.0293 – 0.341
8	0.150	0.0103 – 0.305
9	0.179	0.0312 – 0.342
10	0.146	0.00851 – 0.302
11	0.163	0.0237 – 0.318
12	0.142	0.00884 – 0.291
13	0.192	0.0380 – 0.361
14	0.156	0.0133 – 0.317
15	0.175	0.0292 – 0.334
16	0.149	0.0118 – 0.303
17	0.109	-0.0238 – 0.258
18	0.0839	-0.0413 – 0.226
19	0.0998	-0.0286 – 0.242
20	0.0819	-0.0400 – 0.218
21	0.113	-0.0217 – 0.263
22	0.0841	-0.0407 – 0.226

23	0.102	-0.0267 – 0.244
24	0.0823	-0.0394 – 0.218
25	0.121	-0.0174 – 0.278
26	0.0928	-0.0356 – 0.240
27	0.111	-0.0216 – 0.260
28	0.0912	-0.0337 – 0.233
29	0.121	-0.0167 – 0.276
30	0.0919	-0.0352 – 0.237
31	0.110	-0.0215 – 0.257
32	0.0920	-0.0332 – 0.234
33	0.306	0.108 – 0.510
34	0.373	0.153 – 0.607
35	0.314	0.115 – 0.520
36	0.388	0.164 – 0.624
37	0.295	0.101 – 0.499
38	0.362	0.142 – 0.600
39	0.302	0.108 – 0.505
40	0.365	0.158 – 0.583

Avoiding the Siren Call of the Clock in "Unreasonable Delay" Data Breach Notification Cases

Evan Yahng

Avoiding the Siren Call of the Clock in "Unreasonable Delay" Data Breach Notification Cases

*Evan Yahng**

As online personally identifiable information, data breaches, blockchain, Artificial Intelligence, and other trends in the cyber ecosystem proliferate, courts must confront legal questions about data privacy that past courts have kicked down the road. One such question that courts and scholars have yet to properly interrogate is what constitutes "unreasonable delay" in violation of state data breach notification statutes.

All fifty states, the District of Columbia, Puerto Rico, the Virgin Islands, and Guam have laws requiring companies that hold data to provide notice to data subjects in the event such data is compromised. But courts have been able to punt the question of delay until now because data breach litigation often dies early, either because the statute does not provide a cause of action, or because the harm is too speculative to support a cause of action in negligence. With courts recognizing more statutory causes of action and more harms in negligence, the time to answer the question has come. Indeed, a massive number of courts addressed unreasonably delayed data breach notice claims in 2024.

A substantial number of those courts made a grave error when they denied motions to dismiss solely because precedent in their jurisdictions held that a given number of days was prima facie unreasonable. This Article argues that this approach misunderstands the purpose of data breach notification laws and leads to undesirable results including costly liability for companies and risks to individual consumers' identities. After sampling some of these 2024 cases, this Article explains myriad problems associated with relying on the clock as the sole indicator of reasonableness. Finally, this Article suggests that courts follow a more practical and doctrinally desirable approach whereby they examine defendants' post-breach investigation to determine whether any delay was unreasonable.

---

* J.D., Washington University in St. Louis; B.A., Xavier University. The author is a private attorney. Any and all views and opinions expressed herein belong solely to the author and do not necessarily reflect the views of the author's employer.

I. INTRODUCTION

Everyone has had a data breach notification letter appear in their mailbox, but who is to blame when it comes too late? Your credit card has already been used, a bank account has already been opened in your name, and your inbox has already been filled with spam. Courts across the country have recently confronted whether a defendant's delay in sending notice of a data breach to the individuals whose information was exposed was "unreasonable" or not "as soon as possible" in violation of state law. A number of these courts denied motions to dismiss solely on the basis of precedent that a given number of days' delay was prima facie unreasonable.

This Article warns that relying on case law to determine whether a given number of days' delay is unreasonable misunderstands the purpose of data breach notification laws and leads to undesirable results. Courts should understand that whether a delay in sending breach notification is unreasonable is unripe to be resolved at the motion to dismiss stage. Courts must instead make case-by-case determinations in light of the totality of the circumstances at the motion for summary judgment stage or later. This requires interrogating the defendant's post-breach investigation, which may include retaining counsel, containing the breach, accurately determining the number and identity of victims, drafting the notice, and more. This approach is more thorough, beneficial to businesses and individuals, coherent in the long term, and consistent with the statutory purpose of data breach notification laws.

This Article proceeds in three parts. The first part samples some of the recent case law regarding whether delay in sending data breach notification is unreasonable. The second part explains the pitfalls of using the clock as the only indicator of reasonableness without interrogating the defendant's incident response, and why examining the facts of the post-breach investigation is practically and doctrinally more desirable for courts, litigants, affected individuals, and businesses. Finally, the third part considers the elements of a post-breach investigation that courts should examine when determining reasonableness. It should be noted that this Article focuses on whether a violation of a state data breach notification statute has occurred because the delay was unreasonable. This is separate from whether a delay was negligent, or whether it caused

cognizable injury or harm for standing, negligence, or other purposes. These questions have been well litigated (although courts have not reached a unanimous resolution) and are beyond the scope of this Article.

II. SUMMARY OF RECENT DELAY-OF-BREACH-NOTIFICATION CASES

2024 was a boom year for litigation of data breaches. Data breaches are an increasing problem for businesses and the millions of individuals whose data is in their hands.[1] Data incidents can result from criminal conduct such as hacking, insider theft, or phishing, as well as accidents such as mistaken publication or lost computers.[2] When unavoidable, a company can mitigate the effects of a data incident through actions including prompt containment, eradication, recovery, and notification to affected data subjects.[3] When mishandled, the costs of data incidents, both to businesses and data subjects, can be enormous. In 2024, the global average cost of a data breach was $4.88 million—a ten percent increase over the last year.[4] This problem is even worse in sectors like healthcare.[5] Individuals, meanwhile, may suffer the crippling financial and emotional consequences of identity theft that result from a data incident.

All fifty states, the District of Columbia, Puerto Rico, the Virgin Islands, and Guam have laws requiring companies that do business in their jurisdiction and that hold computerized data to provide notice to affected citizens in the event such data is compromised.[6] Depending on the risk of harm and the number of individuals or citizens affected, some of these laws also require

---

1 140 AM. JUR. *Trials* § 1 (2015); *see also* Evan M. Wooten, *The State of Data-Breach Litigation and Enforcement: Before the 2013 Mega Breaches and Beyond*, 24 J. ANTITRUST & UNFAIR COMPETITION L. 229, 229 (2015) ("Corporate legal spending on data security in the United States increased from $1 billion in 2013 to $1.4 billion in 2014, and is expected to climb to $1.5 billion in 2015—a 7.9% increase that dwarfs the next highest practice area (2.7% for class actions).").

2 140 AM. JUR. *Trials*, *supra* note 1, § 2.

3 *Id.*

4 *The Cost of Data Breaches*, THOMSON REUTERS (Dec. 11, 2024), https://legal.thomsonreuters.com/blog/the-cost-of-data-breaches/ [https://perma.cc/6VCP-Z3Z9].

5 140 AM. JUR. *Trials*, *supra* note 1, § 2 ("The study found the cost of these breaches to health care organizations in 2014 was significantly more expensive than in any other sector of the economy at $359 per capita in the health care sector compared to $206 in the financial services industry and $155 in consumer products organizations.").

6 David Garrison Golubock, *Remote Workers, Ever-Present Risk: Employer Liability for Data Breaches in the Era of Hybrid Workplaces*, 15 CASE W. RESERVE J.L. TECH. & INTERNET 305, 336 (2024).

notice to be given to a regulator and credit reporting agencies.[7] In all fifty states, the state attorney general may impose fines for violations, while approximately fifteen states also grant a private right of action to individuals harmed by the breach.[8] Many federal privacy statutes and regulations also require data breach notification and employ varying frameworks, though there is no general federal reporting law.[9]

While notice requirements vary by jurisdiction, companies must generally provide a description of the incident, including an approximate date; the types of data affected; the remedial steps the company has taken; and information regarding credit freezes, monitoring, and contact information for obtaining assistance from the Federal Trade Commission. While many states express a preference for notification by mail, states often permit alternative methods of notification such as telephone, email, or public posting under certain circumstances, including undue financial burden or insufficient consumer contact information.[10]

One of the main questions a company must answer when a data breach occurs is when to send notice. Some states require notification without unreasonable delay, subject to a specific maximum time limit. For example, Florida's statute reads:

> Notice to individuals shall be made as expeditiously as practicable and without unreasonable delay, taking into account the time necessary to allow the covered entity to determine the scope of the

---

7 *See The Ultimate Guide to Data Breach Notification Laws by State*, EMBROKER (Feb. 28, 2025), https://www.embroker.com/blog/data-breach-laws-by-state/ [https://perma.cc/RPP8-DNMC].

8 *See id.*; *see also* PETER SWIRE AND DEBRAE KENNEDY-MAYO, U.S. PRIVATE-SECTOR PRIVACY LAW AND PRACTICE FOR INFORMATION PRIVACY PROFESSIONALS 325 (4th ed. 2024) ("Nearly 15 states grant a private right of action to individuals harmed by disclosure of their personal information.").

9 *See* Golubock, *supra* note 6, at 337 (summarizing the existence or non-existence of a private right of action to enforce data breach notification requirements under the Health Insurance Portability and Accountability Act, Federal Communications Commission rules, Securities and Exchange Commission rules, Federal Trade Commission rules, Fair Credit Reporting Act, and Cyber Incident Reporting for Critical Infrastructure Act of 2022); Nicole B. Perkins, *Spreading a Digital Disease: The Circuit Split on Data Breaches and Its Effects on the Health Sector*, 20 IND. HEALTH L. REV. 435, 456 (2023) (discussing the Health Information Technology for Economic and Clinical Health Act's requirement that notification be provided without unreasonable delay and in no case later than sixty days following the discovery of a breach).

10 *See, e.g.*, CAL. CIV. CODE § 1798.82(j) (West 2025) (permitting written notice, electronic notice, or substitute notice if the cost of providing notice would exceed $250,000, the class of persons to be notified exceeds 500,000, or if the person or business does not have sufficient contact information).

> breach of security, to identify individuals affected by the breach, and to restore the reasonable integrity of the data system that was breached, but no later than 30 days after the determination of a breach or reason to believe a breach occurred . . . .[11]

The time limit may vary from thirty to ninety days, with an average of forty-five days,[12] though industry best practice for notification has converged on seventy-two hours due to the influence of the European Union's General Data Protection Regulation.[13] Other jurisdictions only require that notice is provided "without unreasonable delay" or "as soon as possible," without providing a defined time limit. For example, Arkansas' statute provides: "The disclosure shall be made in the most expedient time and manner possible and without unreasonable delay, consistent with the legitimate needs of law enforcement . . . or any measures necessary to determine the scope of the breach and to restore the reasonable integrity of the data system."[14] This Article is principally concerned with statutes that follow the latter formulation.[15] While the question of when a delay is unreasonable can be relevant to statutes that set a specific number of days by which notification must be given,[16] there is less need for guidance (and the matter is less

---

11 FLA. STAT. § 501.171(4)(a) (2025); *see also* DEL. CODE ANN. tit. 6, § 12B-102(c) (2025) ("Notice . . . must be made without unreasonable delay but not later than 60 days after determination of the breach of security . . . ."); ALA. CODE § 8-38-5(b) (2025) ("Notice to individuals . . . shall be made as expeditiously as possible and without unreasonable delay . . . . [T]he covered entity shall provide notice within 45 days of the covered entity's receipt of notice from a third-party agent that a breach has occurred . . . .").

12 Carol M. Hayes, *Comparative Analysis of Data Breach Laws: Comprehension, Interpretation, and External Sources of Legislative Text*, 23 LEWIS & CLARK L. REV. 1221, 1260 (2020).

13 Scott J. Shackelford, Anne Boustead & Christos Makridis, *Defining "Reasonable" Cybersecurity: Lessons from the States*, 25 YALE J.L. & TECH. 86, 109–10 (2023).

14 ARK. CODE ANN. § 4-110-105(a)(2) (2025); *see also* GA. CODE ANN. § 10-1-912(a) (2025); ALASKA STAT. § 45.48.010 (2024); IOWA CODE § 715C.2 (2025); MONT. CODE ANN. § 2-6-1503(1)(a)–(b) (2025). Most statutes also permit delay if notification may impede a criminal investigation. *See, e.g.*, WYO. STAT. ANN. § 40-12-502(a)–(b) (2025) (providing both that notice shall be made "in the most expedient time possible and without unreasonable delay, consistent with the legitimate needs of law enforcement and consistent with any measures necessary to determine the scope of the breach and to restore the reasonable integrity of the computerized data system," and permitting notification to be "delayed if a law enforcement agency determines in writing that the notification may seriously impede a criminal investigation").

15 The clock generally begins to run when a company "reasonably" believes, suspects, or confirms a data breach has occurred. When the clock starts ticking is separate from the reasonableness of the delay once the clock has started. *See* Hayes, *supra* note 12, at 1260–61.

16 For example, a court in Florida may need to determine whether a delay of less than thirty days was unreasonable or not as expeditiously as practicable. *See id.* at 1261.

likely to be disputed) when the legislature has established a specific timeline.[17]

Neither courts nor scholars have settled on what constitutes "unreasonable delay" in suits brought by individuals whose data was exposed.[18] This is due to the fact that data breach litigation often dies on the vine early, either because the statute does not provide a cause of action,[19] or because the harm is too speculative to support a cause of action in negligence.[20] But as data breaches continue to proliferate, breach-related lawsuits become more common, and as more courts recognize breach-related harms in negligence,[21] courts will have to address this issue. This Part highlights a sampling of the many data breach lawsuits from 2024 that measured the reasonableness of a delay by comparing it to timelines set by precedent. The sheer number of data breach cases that were brought in 2024 makes it impossible to summarize them all; therefore, this Article is not intended to be an exhaustive list.

A. California

Due to the number of Fortune 500 companies, data brokers, and other cyber-trailblazers in the state, California was unsurprisingly a hotbed for litigation of this issue in 2024. Under California's Customer Records Act (CRA), any person or entity that conducts business in California, and that owns or licenses computerized data containing personal information, must disclose a security breach of their system to all California residents whose unencrypted information (or encrypted information along with the key) was, or is reasonably believed to have been, acquired by an unauthorized person.[22] Notification

---

17 Also in Florida, after thirty days, this analysis would not be necessary. *See id.*

18 *See* Golubock, *supra* note 6 ("In theory, many of these laws do create a private right of action against businesses that fail to timely disclose a data breach. In practice, however, many of these statutes leave it up to courts to determine whether a business delayed unreasonably in notifying affected parties, and even delays of weeks may not be sufficient to support a claim for failure to notify.").

19 *See, e.g.*, N.Y. GEN. BUS. LAW § 899-aa (McKinney 2025); GA. CODE ANN. §§ 10-1-910 to -915 (West 2025).

20 *See, e.g.*, Mohsen v. Veridian Credit Union, 733 F. Supp. 3d 754, 763–65 (N.D. Iowa 2024) (holding that the economic loss rule barred a data breach-related negligence claim).

21 *See, e.g.*, Nunley v. Chelan-Douglas Health Dist., 558 P.3d 513, 517 (Wash. Ct. App. 2024) (holding that loss of time, mental distress, and loss of value of personal information satisfied Washington's common-law requirement that a plaintiff in a negligence case must prove damages).

22 CAL. CIV. CODE § 1798.82(a) (West 2025).

must be provided "in the most expedient time possible and without unreasonable delay," though no specific timeframe for disclosure is mandated.[23]

For example, in *Jackson v. Health Center Partners of Southern California*, the plaintiff alleged that the defendant's 139-day delay in disclosing a data breach violated the CRA.[24] Citing precedent, and without questioning the defendant's breach response, the court wrote:

> Some courts have found that five-month delays and nine-month delays in providing notice of a data breach sufficiently alleged an "unreasonable delay" under the CRA. In contrast, an alleged ten-day delay was not a sufficient allegation of unreasonable delay.
>
> . . . [A] court in a similar case alleging a violation of the CRA denied a motion to dismiss [and] . . . set for trial the CRA claim of a one-month delay . . . .
>
> . . . Today, Plaintiff's allegation of harm is sufficient, along with the allegation of unreasonable delay, to state a plausible state law cause of action under the CRA . . . .[25]

The court thus implicitly rejected the theory that the reasonableness of a particular delay under the CRA is a question for trial rather than for a motion to dismiss.[26]

The *Jackson* court cited *J.M. v. Illuminate Education, Inc.* for the proposition that a five-month delay was sufficient to survive a motion to dismiss.[27] There, the plaintiff alleged that the defendant, an education consultant, delayed disclosure of a breach for five months in violation of the CRA.[28] The California Court of Appeal held that a "five-month disclosure delay supports a cause of action under the CRA because such a delay prevents victims from taking prompt steps to protect their personal information."[29] Citing precedent regarding whether a plaintiff suffered injury, the court stated: "A delay of even three months in

---

23 *Id.*

24 Jackson v. Health Ctr. Partners of S. Cal., No. 24-cv-00106-BEN (DDL), 2024 WL 3708867, at *5 (S.D. Cal. Aug. 7, 2024).

25 *Id.* (citations omitted) (first citing J.M. v. Illuminate Educ., Inc., 323 Cal. Rptr. 3d 605, 612 (Cal. Ct. App. 2024); then citing *In re* Sony Gaming Networks & Customer Data Sec. Breach Litig., 996 F. Supp. 2d 942, 1010 (S.D. Cal. 2014)).

26 *See id.* ("And at least one court has suggested that whether a particular delay qualifies as an 'unreasonable [delay] under' the CRA is normally a question for trial rather than for a motion to dismiss."). *Contra In re Sony*, 996 F. Supp. 2d at 1010.

27 *Jackson*, 2024 WL 3708867, at *5 (citing *J.M.*, 323 Cal. Rptr. 3d at 612).

28 *J.M.*, 323 Cal. Rptr. 3d at 609.

29 *Id.* at 613 (citation omitted).

notifying victims has been held to be sufficient to state a cause of action for damages under the CRA."[30]

B. Florida

On the other side of the country, the U.S. District Court for the Southern District of Florida addressed claims that a supplemental-benefits insurance provider's notice that the plaintiff's personal data was compromised under the CRA was unreasonably delayed.[31] Brushing past the issue, the court wrote: "Plaintiffs' allegation that Defendants failed to disclose the Data Breach 'in a timely and accurate fashion,' waiting until at least April 13, 2023, despite being notified by February 3, 2023, is similarly sufficient."[32] The court reasoned only that, at the pleading stage, it was bound to accept the plaintiffs' allegation that a two-month delay was unreasonable.[33]

C. New England

The Massachusetts District Court reached a similar holding when reviewing a putative subclass alleging a violation of New Hampshire's Notification of Security Breach Required law in its decision in *In re Shields Health Care Group, Inc. Data Breach Litigation*.[34] Plaintiffs, patients of a medical scanning and surgical services company whose data was compromised in a breach, alleged that the defendant took approximately four months to provide notification.[35] As a result, they claimed, they could not take measures to prevent injuries resulting from their information being for sale on the dark web, including fraudulent bank charges, suspicious email activity, emotional distress, and

30 *Id.* (citing *In re* Ambry Genetics Data Breach Litig., 567 F. Supp. 3d 1130, 1150 (C.D. Cal. 2021)).

31 *In re* Fortra File Transfer Software Data Sec. Breach Litig., No. 23-cv-60830-RAR, 2024 WL 4547212, at *14 (S.D. Fla. Sept. 18, 2024). The CRA claim was brought by a putative subclass comprised of California residents. *Id.* at *13.

32 *Id.* at *14 (citations omitted).

33 *Id.* It is not uncommon for courts to dedicate little ink to this issue. For example, in *Dusterhoft v. OneTouchPoint*, a Wisconsin case addressing claims under South Carolina's breach notification statute, the only attention the court gave the unreasonable delay issue was in the following sentences: "[The c]omplaint alleges that OneTouchPoint did not notify [plaintiff] of the breach until . . . a full three months after the breach occurred. This significant delay is sufficient to create a plausible inference that OneTouchPoint failed to notify [plaintiff] 'in the most expedient time possible,' as required by the statute." Dusterhoft v. OneTouchPoint Corp., No. 22-cv-0882-bhl, 2024 WL 4263762, at *20 (E.D. Wis. Sept. 23, 2024).

34 *In re* Shields Health Care Grp., Inc. Data Breach Litig., 721 F. Supp. 3d 152, 167–68 (quoting N.H. REV. STAT. ANN. § 359-C:20(I)(a) (2025)).

35 *Id.* at 158–59.

loss of value of the data.[36] The defendants moved to dismiss the New Hampshire breach notification claim, but the court denied the motion.[37]

New Hampshire's security breach law requires any person doing business in the state "who owns or licenses computerized data that includes personal information [to], when it becomes aware of a security breach, promptly determine" whether a misuse of the information has occurred or is reasonably likely to occur.[38] If the determination is made in the affirmative or if a determination cannot be made, the person must "notify the affected individuals as soon as possible."[39] The *Shields* subclass alleged the defendant's four-month delay was not "as soon as possible" as required by the statute.[40] The defendant rebutted that it immediately launched an investigation, and the complaint did not show that three months was an unreasonable investigation period.[41] The court compared New Hampshire's statute to other state laws that require companies to notify individuals of data breaches "without unreasonable delay," and concluded: "Courts interpreting statutes with similar language have not dismissed claims where the defendant waited nine months, five months, and four months to notify plaintiffs of a data breach. Thus, [plaintiff] has stated a claim under the New Hampshire notice statute."[42]

III. BLINDLY FOLLOWING CASE LAW REGARDING HOW MUCH TIME CONSTITUTES AN UNREASONABLE DELAY IS MISGUIDED

These 2024 cases relied on precedent as the end-all, be-all of what is or is not per se unreasonable for motion to dismiss purposes. In other words, the courts determined that if one hundred days had been held unreasonable in the past, then one hundred days must be prima facie unreasonable in the case before them. While this approach is not new, it is and always has been wrong. Using the clock alone is tempting and, at first glance, fits with how courts generally apply stare decisis. However, this approach misses what is actually going on. It is not that hours, days, or months themselves are unreasonable, but

36 *See id.* at 159.
37 *See id.* at 167–68.
38 N.H. REV. STAT. ANN. § 359-C:20(I)(a).
39 *Id.*
40 *In re Shields*, 721 F. Supp. 3d at 167.
41 *Id.*
42 *Id.* (citations omitted).

rather what caused the delay that may or may not be reasonable. Whether a data breach response is unreasonable cannot be divined by citations to other cases' timelines for three reasons: (1) it is unripe to be determined at the pleading stage, (2) it creates impractical results, and (3) it is inconsistent with the purpose of data breach statutes.

A. Whether a Delay in Sending a Data Breach Notice Is Unreasonable Is a Question of Fact to Be Determined at the Summary Judgment Stage or Later

That a data breach response was unreasonably delayed must, of course, be in the complaint. Therefore, courts can dismiss a case if the plaintiff fails to allege that a delay was unreasonable[43] or if there is a failure to allege a timeline from which it could be inferred that an unreasonable delay occurred[44] because a core element of the claim would be missing. But if a defendant moves to dismiss a properly alleged delay on the ground that the delay was not unreasonable, courts should refuse to decide the issue as unripe at that early stage. In data breach cases, as in most contexts, reasonableness itself is a question of fact for the factfinder to decide after weighing the evidence and credibility of witnesses.[45]

In individual cases, the practical result of either approach is the same: the motion to dismiss is denied. However, the "why" is important at the macro level. Data breaches are fact-intensive disputes that must address the number of individuals affected, the systems compromised, the method of exposure, the injuries the plaintiff suffered, and more. "Whether a delay was reasonable requires courts to look beyond the length of the delay and consider the facts alleged."[46] Inflexible, bright-line rules for

---

43 *See* Razuki v. Caliber Home Loans, Inc., No. 17-cv-1718-LAB (WVG), 2018 WL 6018361, at *2 (S.D. Cal. Nov. 15, 2018) (finding that dismissal of the lawsuit was appropriate because the plaintiff failed to claim that a five-month delay was unreasonable).

44 *See In re* Yahoo! Inc. Customer Data Sec. Breach Litig., 313 F. Supp. 3d 1113, 1146 (N.D. Cal. 2018) ("Without more specific information, the Court cannot evaluate whether Defendants unreasonably delayed in notifying customers about the extent of the 2013 Breach . . . . Plaintiffs' allegations remain insufficient.").

45 *See* Stallone v. Farmers Grp., Inc., No. 2:21-cv-01659-GMN-VCF, 2022 WL 10091489, at *8 (D. Nev. Oct. 15, 2022) ("Defendants' argument [that plaintiffs did not incur injury due to delay] is better suited for a motion for summary judgment when the record is more fully developed."); *see also* Buonasera v. Honest Co., 208 F. Supp. 3d 555, 566 (S.D.N.Y. 2016) ("Courts have generally held that since this second factor requires a reasonableness analysis, it cannot be resolved on a motion to dismiss.").

46 Griffey v. Magellan Health Inc., 562 F. Supp. 3d 34, 56 (D. Ariz. 2021).

unreasonableness at the motion-to-dismiss stage incentivize hasty and sloppy breach notification.[47] A hardline rule that a certain number of days is per se unreasonable pushes businesses to respond as fast as possible, even if they have not properly restored the integrity of the system, determined the scope of the breach, or accurately established data subjects' contact information. If courts wait until they have a record to determine what is reasonable, businesses have a reason to take robust incident response measures, which incentivizes a prudent breach response. The court's review of which measures were reasonably necessary prevents reporting entities from delaying notice in bad faith. Waiting until summary judgment for the right reasons makes breach responses more accurate and more precise.

Notwithstanding the aforementioned 2024 decisions, much of the case law supports this position. For example, the U.S. District Court for the Eastern District of Virginia has opined:

> [W]hether [defendant's] substitute notice was timely is a question not ripe for the motion to dismiss stage. The notice's timeliness is a factual question that asks whether notice of the data breach occurred "without unreasonable delay." Here, the Complaint alleges that it took [defendant] approximately four . . . months to realize that there had been a breach, which, in fact, [defendant] did not itself discover. Moreover, the Complaint alleges that [defendant] could have discovered the hack as early as April, since the hacker . . . had posted her action on an online forum . . . . Further, there is no argument by [defendant] that the law enforcement safe harbor, which permits a delay in notifying affected individuals, applies. In

---

47 An upper limit by which notification must be given "would likely incentivize [companies] to notify individuals quicker than they otherwise would." Michael Bloom, *Protecting Personal Data: A Model Data Security and Breach Notification Statute*, 92 ST. JOHN'S L. REV. 977, 997 (2018).

> But, bright line rules are inflexible. . . . [A]n upper limit may do more harm than good. There may be situations where an entity has the means to notify individuals in much less time than the commonly required thirty days. Including an upper limit on what can be considered "without undue delay" can actually give entities a "cushion to delay notification[]." Some businesses argue that thirty days is too short of a window to assess the extent of and respond to a data breach. In that event, when that claim is true and stands up to scrutiny from federal agencies, a more flexible window would allow entities to delay notification until it would be more proper. As long as it is objectively reasonable that the entities take that much time, it would be fairer to allow them to do so. The uncapped standard provides flexibility to deal with the exigencies of each individualized situation and is the preferable standard for a federal data breach notification law.

Id. (second alteration in original) (footnotes omitted).

> any event, these are all factual questions not suitable for disposition on a motion to dismiss.[48]

Likewise, the Nebraska District Court declined to rule on whether the plaintiff could establish that the delay was unreasonable because it presented a question that went "to the sufficiency of the evidence supporting the allegations in the amended complaint, not the sufficiency of the allegations."[49] Even in California, where some courts have held to the contrary, many have held that the reasonableness of breach notice is "a factual determination not properly decided by the Court on a motion to dismiss."[50]

B. Relying on Precedent to Determine the Number of Days that Is Prima Facie Unreasonable Leads to Impractical Results

If courts follow the "*x* number of days is per se a sufficient allegation of unreasonableness" approach, then results within a state would become incoherent and unjust. For example, if California followed the *Jackson* court's logic, where an allegation of a 139-day delay alone is sufficient to survive a motion to dismiss on its own terms, then the most responsible California company, which conducted a rigorous investigation lasting 140 days in good faith, loses simply because more than 139 days is per se unreasonable. This would be the case even if no one could possibly have completed the investigation sooner.[51] Turn the hypothetical on its head and you get equally ridiculous results. If courts looking exclusively at the timelines hold that 139 days is per se unreasonable, then it is possible to imagine a company that sits on its hands for 138 days while conducting a meager or no investigation, only to send notice at the eleventh hour. Clearly, this result is not desirable, and this delay is not reasonable.

---

48 *In re* Cap. One Consumer Data Sec. Breach Litig., 488 F. Supp. 3d 374, 416 (E.D. Va. 2020) (citations omitted).

49 Weisenberger v. Ameritas Mut. Holding Co., 597 F. Supp. 3d 1351, 1365 (D. Neb. 2022) (citing Stamm v. County of Cheyenne, 326 F. Supp. 3d 832, 847 (D. Neb. 2018)).

50 *See In re* Sony Gaming Networks & Customer Data Sec. Breach Litig., 996 F. Supp. 2d 942, 1009 (S.D. Cal. 2014).

51 Of course, this would be an extraordinary situation, and in the vast majority of cases, four-month delays would be difficult to justify. This Article does not claim that a data breach response should take more than a few days or weeks, but rather, it clarifies how courts should determine the appropriate length of a data breach response. *But see* Golubock, *supra* note 6, at 343 ("[C]ourts have permitted reporting of data breaches weeks after the facts of a breach became known.").

The results would be no more comprehensible across state lines. Imagine a hypothetical State *A*, which has case law that only a 45-day or more delay is per se unreasonable. Also imagine a hypothetical State *B*, which has case law that only a 60-day or more delay is per se unreasonable. A 46-day delay would be prima facie unreasonable in State *A* but prima facie reasonable in State *B*. Is there some public policy specific to State *B* that demands an extra 15 days in the grace period? Perhaps, if all the companies in State *B* are so complex and hold data on so many individuals that a longer grace period is desirable. But this seems far-fetched, as does the idea that a 60-day grace period—as opposed to a 45-day grace period—would be top-of-mind for State *B* voters in electing their legislators.

This inconsistency is not just doctrinal—it has real-world harms. First, staying with the State *A*/State *B* hypothetical, companies that participate in interstate commerce would have to follow arbitrarily different timelines for the same investigation, increasing compliance costs and potentially delaying notification for no apparent benefit. This exacerbates the fact that data breaches can lead to astronomically costly litigation. Professors Daniel Solove and Danielle Keats Citron have identified what they call the "multiplier problem," noting that organizations hold data on so many individuals that recognizing even a small amount of harm is multiplied by a staggering number of people such that runaway class actions could bankrupt companies.[52] And for what purpose? The relief provided by slow and expensive class action lawsuits is unlikely to provide meaningful redress to the vast majority of data subjects.[53] Many business advocates, therefore, argue that letting data breach liability off the leash would bankrupt any business, small or large, that holds data. Because unlawful disclosures impact "tens of thousands of individuals[,] . . . [t]he liability faced by an allegedly negligent defendant would be catastrophic in magnitude."[54]

This is not necessarily to say liability for data breaches should be wholly off the table. Rather, this shows that courts

---

52 Daniel J. Solove & Danielle Keats Citron, *Risk and Anxiety: A Theory of Data-Breach Harms*, 96 TEX. L. REV. 737, 783 (2018).

53 *Id.*

54 Kenneth S. Abraham & G. Edward White, *Torts Without Names, New Torts, and the Future of Liability for Intangible Harm*, 68 AM. U.L. REV. 2089, 2137 (2019).

must approach data breach liability with extreme caution. Herein lies the proper role of stare decisis in data breach response law: determining which elements of an incident response investigation are reasonable for breaches of which scope and in which sectors. Transparency in what breach response steps are reasonable sets clear guidelines for fostering customer confidence and avoiding liability. Take the example of the hypothetical high-end cyber forensics group, Data Breach Responders Pro, which takes a meticulous—if sometimes slow—approach to cyber incidents. Assume a prior court determined that it was reasonable for Company *X*, a fictional company that suffered a data breach, to retain Data Breach Responders Pro to contain and analyze a breach of one million individuals' protected records. Company *Y*, another fictional company that suffered a data breach of one million individuals' similar records, can confidently retain Data Breach Responders Pro to contain and analyze its own breach without worrying that Data Breach Responders Pro's thorough process will unreasonably delay notice in violation of the law. Likewise, if the case law indicates that certain kinds of data, such as biometric data or Social Security numbers (SSNs), justified a longer period of analysis in Company *X*'s case, then Company *Y* can be equally thorough when dealing with biometric data or SSNs. On the flip side, if an earlier court held that it was unreasonable for Company *X* to take extra time drafting an extensive data breach notification letter given the low risk of harm involved, then Company *Y* knows not to do so.

This approach helps consumers as well. From the individual's perspective, data breach notices are important so that consumers can activate credit monitoring services and take other steps to mitigate the effects of a breach. A substantively inaccurate data breach notification (for instance, what data was compromised, how it was compromised, to whom data was disclosed, and what risks may be present) may mean that individuals retain the wrong kind or degree of identity protection. An underinclusive data breach notification could leave some individuals who should be entitled to protection on the company's dime without any identity protection at all. And an overinclusive data breach notification may cause the company to lose money paying for individuals who have no need to retain identity protection services, while those individuals take time out of their busy lives and endure the emotional stress of suffering identity theft for no reason. Timely, but also thorough and

accurate notice of a data breach, is in both the business' and the consumers' best interest.

C. Ignoring Breach Response Investigations in Favor of a Topline Number of Days Is Inconsistent with the Purpose of Data Breach Response Statutes

Data breach notification statutes do not permit delay for delay's sake but rather so that companies have time to determine root causes and take responsible steps toward remediation. Therefore, a peek under the hood is most consistent with the statutory purpose of data breach laws. Some data breach notification statutes explicitly tie reasonableness to the investigation. For example, Alabama's statute states that "[n]otice to individuals . . . shall be made as expeditiously as possible and without unreasonable delay, taking into account the time necessary to allow the covered entity to conduct an investigation."[55] In states like Alabama, the importance of the investigation is obvious. Yet, as previously discussed, other states favor an approach that simply requires notification to be provided "without unreasonable delay."[56] This approach, too, serves as an implicit authorization for investigation if none exists in the statute and, by extension, a built-in explanation of what delay is unreasonable.[57] Researcher Carol Hayes has argued that the phrase "without unreasonable delay" is preferable to phrases like "as quickly as possible" or "as soon as possible" because it "allows for reasonableness considerations to be a factor in enforcement."[58] She explains:

> The focus on unreasonable delays implies that there could be a reasonable delay. Forty-two of the analyzed laws include language suggesting that a reasonable delay would include time to recover from the breach. This is commonly phrased to include time to determine the scope of the breach and time to restore system integrity. All of the analyzed data breach laws included explicit language allowing for delays due to a law enforcement investigation related to the breach.[59]

---

55 ALA. CODE § 8-38-5(b) (2025).

56 *See supra* Part II.

57 Dana J. Lesemann, *It's Not the Breach, It's the Cover-Up: Using Digital Forensics to Mitigate Losses and Comply with Florida's Data Breach Notification Statute*, 82 FLA. BAR J. 20, 24 (2008).

58 Hayes, *supra* note 12.

59 *Id.* (footnote omitted).

Investigations are so core to breach response that in some states, failure to conduct a good faith investigation is itself a violation. Kansas' notification statute reads:

> A person that conducts business in this state . . . that owns or licenses computerized data that includes personal information shall, when it becomes aware of any breach of the security of the system, conduct in good faith a reasonable and prompt investigation to determine the likelihood that personal information has been or will be misused.[60]

Regardless of how it is styled, all of these approaches center the "who, what, when, where, and how" of the breach response and investigation. Time and breach response go hand in hand. For courts to get the issue right, and to effectuate the purpose of the statutes, they must consider the defendant's investigation.

IV. COURTS MUST INTERROGATE THE ELEMENTS OF A DEFENDANT'S DATA INCIDENT RESPONSE TO DETERMINE WHETHER DELAY IS UNREASONABLE

If courts cannot rely on precedent to determine how many days is per se reasonable, how should they answer the question? Courts must do the sometimes painstaking work of considering the facts of the defendant's response to the breach. Among other factors, this may involve determining: the nature of the information compromised (for instance, comparing protected health information to a date of birth, zip code, or SSN); who gained unauthorized access to the data; the risk of harm presented by disclosure; the number of individuals affected; the availability and accuracy of those individuals' contact information; whether or not the defendant retained counsel; the identity, capabilities, experience, and financial cost of counsel; which systems were affected; how those systems were affected; the complexity of those systems; the time required to restore the integrity of such systems; the level of detail required in the notice to satisfy the reporting requirement; and more. Many state statutes explicitly list what factors are to be analyzed when determining the reasonableness of the response. For example, Hawaii's data breach notification statute requires notification to be made "without unreasonable delay, consistent with . . . any measures necessary to determine sufficient contact information, determine the scope of the breach, and restore the reasonable

60 KAN. STAT. ANN. § 50-7a02 (2025).

integrity, security, and confidentiality of the data system."[61] Courts in states like Hawaii have an easy place to start. Courts in states whose statutes do not enumerate factors can borrow from this guidance in conducting more informed inquiries.

Much of the scholarship has endorsed this approach. Hayes has endorsed a flexible "without unreasonable delay" standard, with the important modification that states clarify which causes for delay are reasonable.[62] She suggests a three-part reasonableness standard for data breach notification delays:

> First, a delay is reasonable if it is necessary for law enforcement purposes. . . .
>
> The second and third parts of the reasonableness standard focus on the data collector's investigation and system restoration. A delay should be considered reasonable if it is necessary to determine the scope of a data breach. This is important because the scope determination is central to data breach notification obligations. A delay should also be considered reasonable if it is necessary to restore the integrity of the affected system. It is important to include recovery time within a reasonableness standard because unless system integrity is restored, a data breach cannot truly be said to be "over."[63]

Likewise, Professors Scott L. Shackelford, Anne Boustead, and Christos Makridis have advocated for "an empirically grounded, flexible approach" to cybersecurity that prioritizes combining cybersecurity best practices and efforts to inform consumers of their rights and the importance of exercising them.[64] Finally, this approach is no harm, no foul for the affected individuals. A court's reasonableness review of the facts will make it unlikely that companies will delay sending notices any more than under the current regime. As previously discussed, for individual litigants, the immediate practical result is the same: denial of a motion to dismiss and resolution of the issue at summary judgment or later.

V. CONCLUSION

Without a general federal data incident reporting statute, making progress in breach notification law is arduous. The decentralized nature of data breach law requires courts and regulators in over fifty jurisdictions to agree on a complex and

---

61 HAW. REV. STAT. § 487N-2 (2025).
62 Hayes, *supra* note 12, at 1276.
63 *Id.* at 1276–77.
64 Shackelford, Boustead & Makridis, *supra* note 13, at 90.

nuanced approach. That the approach explained in this Article asks courts to volunteer for sometimes grueling case-by-case factual analysis only makes it a harder pitch. However, a better approach to what constitutes an "unreasonable" delay in data breach notification statutes—one that focuses on a reporting entity's breach response instead of a topline number of days—is needed to fulfill the purpose of the laws, help businesses maintain compliance, and protect consumers from identity theft.

Fingerprints of Injustice: The Truth Behind Artificial Intelligence and Algorithmic-Driven Evidence

Aubrey A. Butler

Fingerprints of Injustice: The Truth Behind Artificial Intelligence and Algorithmic-Driven Evidence

*Aubrey A. Butler**

Artificial intelligence (AI) has steadily grown in prominence, reaching into nearly every aspect of daily life, and the legal system is not immune from its influence. As various forms of machine evidence have been admitted in court, there has been an accompanying rise in concerns from researchers, judges, and defense attorneys as to the trustworthiness and reliability of this new category of evidence.

Though AI and machine evidence takes various forms, this Note focuses on some of the most prominent programs being used in courtrooms today, namely the probabilistic genotyping software TrueAllele and the recidivism rate prediction software COMPAS. The creators of these programs guarantee their accuracy, yet several studies have demonstrated proven defects in their operation—defects which cannot be fully tested and resolved due to the proprietary or "black box" nature of the underlying source code.

Several solutions have been suggested for how to approach machine evidence moving forward, but this Note posits a new mechanism which has not been previously addressed: the creation of a new federal agency focused on AI within the United States with a department wholly dedicated to computer-driven evidence in the legal system. This agency would be able to analyze machine evidence and send out scientific advisors to courts to counsel judges about the potential dangers of this form of evidence in a way that is not possible under the current system.

* Aubrey Butler graduated with honors from the University of California, Los Angeles with a Bachelor of Science in biology. She is currently a student at the Dale E. Fowler School of Law at Chapman University in California where she is a recipient of both a scholarship and the Distinguished Student Fellowship. She is also a Managing Editor of the prestigious *Chapman Law Review*. She has written extensively inside and outside the classroom with publications in the genres of both fiction and nonfiction.

I. INTRODUCTION

The world is ever changing with new ways to create an easier and more streamlined existence that can meet society's growing needs; but progress and justice are not always aligned. A 2021 survey showed that thirty-seven percent of Americans are worried about the growth of artificial intelligence (AI), particularly as it relates to replacing jobs,[1] but its implications in the legal field are equally concerning. Predictive algorithms, large language models, and probabilistic genotyping—often collectively referred to as types of AI—have already made their way into the court system in a development one judge described as "the beginning to a disastrous end."[2] This disastrous end being the replacement of human judgment by the supposedly objective assessments of computer programs with proven defects.

Cybergenetics' TrueAllele software has identified defendants via mixed DNA in nearly one thousand criminal cases,[3] predictive risk assessment algorithms have aided judges in assessing individuals in the criminal justice system since 1998,[4] and generative AI has already gained popularity as a tool for legal research.[5] Yet, TrueAllele faces constant scrutiny for hiding its source code, assessment software such as Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) was found to be inaccurate twenty-nine percent of the time,[6] and generative AI lacks the necessary safeguards to filter false or misleading information.[7] When life and liberty are at stake, such concerns cannot—and should not—be ignored.

---

1 LEE RAINIE ET AL., AI AND HUMAN ENHANCEMENT: AMERICANS' OPENNESS IS TEMPERED BY A RANGE OF CONCERNS 22 (2022), https://www.pewresearch.org/wp-content/uploads/sites/20/2022/03/PS_2022.03.17_AI-HE_REPORT.pdf [https://perma.cc/3KYS-L24J].

2 Ed Cohen, *Most Judges Haven't Tried ChatGPT, and They Aren't Impressed*, THE NAT'L JUD. COLL. (July 21, 2023), https://www.judges.org/news-and-info/most-judges-havent-tried-chatgpt-and-they-arent-impressed/ [https://perma.cc/96HJ-Y5PV].

3 Justin Jouvenal, *A Secret Algorithm Is Transforming DNA Evidence. This Defendant Could Be the First to Scrutinize It.*, WASH. POST (July 13, 2021), https://www.washingtonpost.com/local/legal-issues/trueallele-software-dna-courts/2021/07/12/66d27c44-6c9d-11eb-9f80-3d7646ce1bc0_story.html [https://perma.cc/X62G-9JCX].

4 *See* Julia Dressel & Hany Farid, *The Dangers of Risk Prediction in the Criminal Justice System*, MIT CASE STUD. IN SOC. & ETHICAL RESPS. COMPUTING, Feb. 5, 2021, at 1, 3.

5 *See* Bernice Bouie Donald et al., *Generative AI and Courts: How Are They Getting Along?*, PLI CHRON., Sept. 2023, at 1, 4.

6 *See* Julia Angwin et al., *Machine Bias*, PROPUBLICA (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [https://perma.cc/52YN-M7G7].

7 *See* Tiernan Ray, *Generative AI Can't Find Its Own Errors. Do We Need Better Prompts?*, ZDNET (Oct. 31, 2023, 8:01 AM), https://www.zdnet.com/article/generative-ai-cant-find-its-own-errors-do-we-need-better-prompts/ [https://perma.cc/F4K2-G4MV].

With the pervasiveness of AI and algorithmic evidence, it would be largely impractical and shortsighted to reject their merits entirely, but measures must be put into place to mitigate the inherent risks. At the forefront of the issue is the fact that companies generally refuse to release their proprietary software's coding, leaving attorneys and judges blind as to how the AI reached a certain conclusion or analyzed the raw data.[8] Despite the lack of transparency, courts have consistently allowed this type of black box evidence to be admitted without requesting that companies release their source code for examination.[9] Among these cases are familiar refrains from companies arguing that keeping trade secrets is necessary to protect business[10] and that defense teams would struggle to decipher the complicated source codes even if they were released.[11]

Recognizing the need to address this growing problem, Chief Justice Roberts made AI the focus of his 2023 year-end report.[12] He warned of AI's various shortcomings, including the possibility of embedded biases, potential violations of due process, and a general lack of reliability, especially pertaining to "hallucinations," the phenomenon where generative AI produces fabricated information.[13] "Machines cannot fully replace key actors in court," the Chief Justice proclaimed.[14] "Nuance matters," and "legal determinations often involve gray areas that still require the application of human judgment. . . . AI is based largely on existing information, which can inform but not make such decisions."[15]

Several solutions have been proposed to overcome this growing conflict, and while many champion amending current

---

8 *See* Katherine Kwong, *The Algorithm Says You Did It: The Use of Black Box Algorithms to Analyze Complex DNA Evidence*, 31 HARV. J.L. & TECH. 275, 287–88 (2017).

9 *Id.* at 284–87, 298; Linton Mann III & William T. Russell Jr., *Disclosure of Software Source Code Not Required to Establish Acceptance of DNA Evidence*, LAW.COM: N.Y. L.J. (May 17, 2022, 12:00 PM), https://www.law.com/newyorklawjournal/2022/05/17/disclosure-of-software-source-code-not-required-to-establish-acceptance-of-dna-evidence/ [https://perma.cc/XVC6-DYSM] (discussing the court case that ruled disclosure of TrueAllele source code was unnecessary because the algorithm was generally accepted in the relevant scientific community).

10 *See* Kwong, *supra* note 8, at 293.

11 *See* Jouvenal, *supra* note 3.

12 *See* JOHN G. ROBERTS, JR., 2023 YEAR-END REPORT ON THE FEDERAL JUDICIARY 2, 5–6 (2023), https://www.supremecourt.gov/publicinfo/year-end/2023year-endreport.pdf [https://perma.cc/YY3T-V68E].

13 *Id.*

14 *Id.* at 6.

15 *Id.*

admissibility rules to allow access to source codes, even if that programming were released to lawyers, they would likely be unable to understand such highly technical information outside their expertise.[16] Others argue that the current rules are more than adequate to evaluate modern evidence, and at most, there must simply be new ways of interpreting those rules when AI is at issue.[17] But in the end, maintaining the status quo would be ineffectual at solving problems that were not in existence at the time the rules were developed.

Instead, this Note offers a new solution which has yet to be fully addressed in any existing literature: a federal agency dedicated to tackling the rising incidence of AI with a division wholly focused on machine evidence in the legal system. This division would review AI and machine-driven evidence being offered in court on a specific case and send a court-appointed scientific advisor trained on the topic to advise the judge. The advisor would be able to better understand complicated source code, be available to guide judges about issues they are not able to research on their own, and would remain subject to confidentiality, solving worries about proprietary software codes being released. Furthermore, having a centralized agency in charge would allow a more intensive and collaborative vetting process when it comes to AI evidence.

Part II begins this discussion with a brief overview of the history of AI, including the distinctions among terms such as "machine learning" and "large language models," and how algorithmic evidence steadily made its way into the court system. Part III takes a deeper look at the problems underlying this type of evidence, from potential violations of the Confrontation Clause to outright false conclusions. Part IV surveys the myriad of solutions that have been posited regarding AI and critically analyzes why each fails to satisfy every facet of this complicated issue. Finally, Part V gives an in-depth look at this new proposition, addressing both the logistics and the fact that it is not an entirely novel idea, but an existing mechanism that can

16 *See infra* Section IV.A; *see also* David A. Prange & Benjamen C. Linden, *Explaining the Almost Unexplainable: Preparing and Presenting Source Code Evidence at Trial*, LAW.COM (June 4, 2020, 10:00 AM), https://www.law.com/legaltechnews/2020/06/04/explaining-the-almost-unexplainable-preparing-and-presenting-source-code-evidence-at-trial/ [https://perma.cc/YAK5-HYQK] (explaining that source code evidence presents "a significant challenge" because of its "obtuse and difficult" nature, the "sheer volume of [which] . . . may require retention of a separate expert" by counsel).

17 *See infra* Section IV.B.

simply be tweaked and applied to the context of AI. Part VI briefly concludes.

II. BACKGROUND

A. The Inception and Limitations of AI

"AI" is an omnipresent fixture of the modern world, two letters blazed across headlines and billboards that have inescapably become part of society's vernacular—two letters with origins far more humble and hopeful than the often sensationalized concerns of "AI takeover"[18] might lead one to believe. In fact, at its inception, AI was never intended to replace human thought or ability, only to enhance the efficiency of fundamentally objective tasks.

Long praised as the father of modern computer science, Alan Turing laid the groundwork for the possibility of AI in 1950 when he wrote a revolutionary paper based upon the simple premise: "Can machines think?"[19] In answering this question, he developed the "imitation game," now known as the Turing Test, which asks whether a blind interrogator evaluating a conversation between a human and a machine could correctly determine which of the two was the person.[20] If the results are indistinguishable and the interrogator cannot correctly choose, the computer can "think," and it passes the test.[21]

At the time Turing published his paper, no machine could come close to winning the imitation game due to a significant limitation in computing ability—"they couldn't store commands, only execute them."[22] A hallmark of human intellect is the ability to learn from mistakes, which is only possible if one remembers they made a mistake in the first place. When a computer cannot store commands, it cannot learn from past behavior, and its intelligence is accordingly restricted.[23] Undeterred by the obstacles surrounding this novel field, computer scientists

---

18 Conor Friedersdorf, *Is This the Start of an AI Takeover?*, THE ATLANTIC (Jan. 3, 2023), https://www.theatlantic.com/newsletters/archive/2023/01/is-this-the-start-of-an-ai-takeover/672628/ [https://perma.cc/8748-SQSH].

19 A.M. Turing, *Computing Machinery and Intelligence*, 59 MIND 433, 433 (1950).

20 *See id.* at 433–34.

21 *See id.*

22 Rockwell Anyoha, *The History of Artificial Intelligence*, HARV. UNIV.: SCI. IN THE NEWS (Aug. 28, 2017), https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/ [https://perma.cc/L2ZW-FEKM].

23 *See id.*

embraced the challenge of creating a thinking machine with fervor, and just five years later, the world's first AI was born.[24]

Stored on punch cards, reliant on the heuristic approach of trial and error, and still debated for its title as "the first AI program," Logic Theorist was programmed to simulate human problem-solving and was able to prove thirty-eight out of the fifty-two mathematical theorems described in *Principia Mathematica*.[25] Logic Theorist was a breakthrough for the time because "it was the first program in symbolic AI, which uses symbols or concepts, rather than data, to train AI to think like a person."[26] As technology advanced and computers became faster and capable of storing greater amounts of data, AI also progressed, and in 1997, IBM's program Deep Blue became the first computer to beat the world's then-reigning chess champion in a match.[27]

As of today, AI has surged into almost every sector through advancements in "natural language processing, image recognition, and automation," and AI adoption by major companies has increased forty-seven percent since 2018.[28] But despite its now overwhelming presence in society, AI is still relatively misunderstood by the general public, and every computer scientist has a different idea of how to define it.[29]

At its heart, AI is a sequence of ones and zeroes.[30] While humans solve problems using abstract thought, machines follow commands written in the only language they can read—binary. The process of doing so is fairly straightforward: a programmer writes an instruction using a programming language, that instruction is translated into binary code (also called machine

---

24 *See id.*

25 Sarah Sloat, *The First AI Started a 70-Year Debate*, POPULAR SCI. (Oct. 3, 2023), https://www.popsci.com/technology/the-first-ai-logic-theorist/ [https://perma.cc/S9DC-FYVY].

26 *Id.*

27 Anyoha, *supra* note 22. However, while Deep Blue managed to win the match, its victory was a result of its speed, not its smarts, as no computer has ever been able to exhibit "humanlike intelligence." *See* Eric Holloway, *For Computers, Smart Is Not the Same Thing as Fast*, MIND MATTERS (Mar. 23, 2021), https://mindmatters.ai/2021/03/for-computers-smart-is-not-the-same-thing-as-fast/ [https://perma.cc/4TW4-7N6X].

28 *The Current Status of Artificial Intelligence*, ALLTECH MAG., https://alltechmagazine.com/what-is/current-status-of-artificial-intelligence/ [https://perma.cc/XA58-VJL3] (July 29, 2024, 12:47 PM).

29 *See id.*

30 *See* Ian Buckley, *What Is Coding and How Does It Work?*, MUO, https://www.makeuseof.com/tag/what-is-coding/ [https://perma.cc/Z5Q4-TPKQ] (June 3, 2021).

code), and the computer follows the instruction.[31] Coding is simply a way of telling the computer what to do; the more complex the command, the more lines of code.

Despite its complicated-sounding name, "algorithm" is just a broad term used to describe a set of instructions for solving a particular problem or executing a particular command.[32] Initial data is input into the algorithm, filtered through the set of directions—which generally take the form of mathematical formulas and problem-solving processes—and final output data is expressed.[33] There are multiple types of algorithms for different applications, but a common, recognizable example is search algorithms, which take input data in the form of key words, search relevant databases for those words, and return the results.[34] Because of the ability to adapt algorithms to perform a variety of functions, they remain vital building blocks of software programs and AI, no matter which form they take.

Terms such as machine learning (ML) and large language model (LLM) are sometimes used interchangeably with AI, but each represents distinct ideas in computer science, and each has inherently different risks and drawbacks. ML represents a subset of AI which uses data and algorithms "to imitate the way that humans learn," gradually improving its accuracy.[35] Unlike algorithms that are used to execute simple commands, ML algorithms are designed to make predictions about patterns of data, determine the error rate of that prediction, and use a model optimization process to better the program's ability to correctly predict outcomes.[36]

ML algorithms are generally trained by inputting data that already have a known output to judge their predictive accuracy, and in this way, the program "uses statistical techniques to help it 'learn' how to get progressively better at a task, without necessarily having been programmed for that certain task."[37] ML

---

31 *Id.*

32 Alexander S. Gillis, *Definition: What Is an Algorithm?*, TECHTARGET, https://www.techtarget.com/whatis/definition/algorithm [https://perma.cc/3WU3-3DES] (last visited Feb. 18, 2025).

33 *See id.*

34 *See id.*

35 *What Is Machine Learning?*, IBM, https://www.ibm.com/topics/machine-learning [https://perma.cc/64P2-RH2P] (last visited May 4, 2025).

36 *Id.*

37 Ellen Glover, *What Is Artificial Intelligence (AI)?*, BUILT IN, https://builtin.com/artificial-intelligence [https://perma.cc/KS4R-98F3] (Dec. 3, 2024).

differs from traditional programming in that it is capable of solving more complex problems, such as recognizing faces or making predictions, and it is trained on sets of data rather than simply running code line by line.[38] However, ML systems still involve "input data . . . fed to an algorithm" just as traditional programming involves algorithms built on top of one another.[39] This means ML accuracy is partly reliant on the initial code or model used to teach the software how to recognize patterns, even if its output is ultimately less predictable than with traditional programming.[40] What actually qualifies as good accuracy for ML programs is highly subjective, and the industry standard is set at a success rate of seventy percent.[41] Seventy percent might be a triumph for the coders who developed the program, but when placed into the context of the justice system, that inaccuracy rate of thirty percent becomes concerning.

Large language models do not seem to fare any better when it comes to accuracy. LLMs are deep learning algorithms, a sub-type of ML that uses an artificial neural network meant to simulate how the brain links disparate ideas.[42] They are designed to process natural language inputs, analyze the patterns and connections between words, and predict how certain sentences will end.[43] The model is trained through exposure to written language via books and articles, and as it analyzes the data, it absorbs grammar, facts, and sentence structure to the point that it can mimic human expression.[44] But this mimicry is not perfect. The LLM can learn biases present in the data it is exposed to, hallucinate information, and its reliability is often

---

38 *See Traditional Programming vs Machine Learning*, INSIGHTSOFTWARE (Feb. 15, 2023), https://insightsoftware.com/blog/machine-learning-vs-traditional-programming [https://perma.cc/S3RQ-UN75].

39 *Id.*

40 *Id.*; *see also* Sara Brown, *Machine Learning, Explained*, MIT SLOAN SCH. OF MGMT. (Apr. 21, 2021), https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained [https://perma.cc/7USN-BAPA] (explaining that "programmers choose a machine learning model to use, supply the data, and let the computer model train itself to find patterns or make predictions," but the programmer still maintains some control over the process as they "can also tweak the model, including changing its parameters" to produce more accurate results).

41 Kirsten Barkved, *How to Know if Your Machine Learning Model Has Good Performance*, ZAMS, https://www.zams.com/blog/machine-learning-model-performance [https://perma.cc/VK9Q-NNKV] (last visited Feb. 18, 2025).

42 *See What Is Machine Learning?*, *supra* note 35.

43 Mike Priest, *Large Language Models Explained*, BOOST.AI, https://boost.ai/blog/llms-large-language-models/ [https://perma.cc/45A2-BPWX] (Feb. 20, 2024).

44 *Id.*

called into question because "due to the innate[ly] unpredictable nature of these models, achieving absolute . . . accuracy is presently unattainable."[45] In theory, ML should catch and fix mistakes, but "current LLMs struggle to self-correct their reasoning," and "expecting these models to inherently recognize and rectify their reasoning mistakes is overly optimistic."[46]

Computers may have seen a rapid evolution over the decades, but this advancement is not indicative of scientists' ultimate triumph over the ones and zeroes nor of the creation of true "artificial intelligence." Several programmers have gone so far as to claim that their AI can successfully pass the Turing Test, but those claims have been widely challenged, in part due to inconsistencies in the test's administration.[47] While "[c]urrent AI systems excel in narrow domains," they "lack the ability to transfer knowledge and skills across different areas of expertise, a hallmark of human intelligence."[48] The stark limitations of AI are only increasingly coming to the forefront as even the largest technology companies have admitted an overall lack of meaningful progress in creating a truly intelligent system.[49] A research team at Apple concluded that "current AI models are 'not capable of genuine logical reasoning,'" and the problem is not one that lends itself to an easy solution, assuming there is any solution at all.[50] The team warns that these concerns should give people caution "as more and more trust is given to AI's 'intelligence,'" which often "isn't what it might appear."[51]

---

45 *Id.*

46 JIE HUANG ET AL., LARGE LANGUAGE MODELS CANNOT SELF-CORRECT REASONING YET 8 (2024).

47 Sanksshep Mahendra, *Has Any AI Passed the Turing Test?*, A.I. PLUS, https://www.aiplusinfo.com/blog/has-any-ai-passed-the-turing-test/ [https://perma.cc/NBA2-9U63] (Jan. 30, 2024, 8:31 PM). Though it remains the most widely used benchmark in evaluating computer intelligence, many criticize the Turing Test's lack of standardized rules, the fact that it does not encompass all forms of human intelligence, and that it could be "passed by unintelligent machines that use tricks and deception to fool humans." *Id.*

48 *Id.*

49 *See* Matthias Bastian, *Truly Intelligent AI: Three Things Google's AI Chief Says Are Missing*, THE DECODER (Jan. 13, 2022), https://the-decoder.com/truly-intelligent-ai-three-things-googles-ai-chief-says-are-missing/ [https://perma.cc/XB52-ECJ9]; Ryan Christoffel, *Apple Researchers Ran an AI Test that Exposed a Fundamental 'Intelligence' Flaw*, 9 to 5 MAC (Nov. 1, 2024, 7:42 AM), https://9to5mac.com/2024/11/01/apple-researchers-ran-an-ai-test-that-exposed-a-fundamental-intelligence-flaw/ [https://perma.cc/58JA-NZ3N].

50 Christoffel, *supra* note 49.

51 *Id.* (describing a test in which AI could not solve a simple math problem when written in word form rather than with pure numbers and when clearly irrelevant information was included in the problem); *see also* Tim Hardwick, *AI Companies*

Concerns of AI takeover create an inflated and overstated idea of AI's true capabilities, painting a picture of computers that can think just as well, or even better, than humans. In reality, computers remain incapable of independent thought, of doing anything beyond following the instructions of a programmer, and AI software is far from being as objective and infallible as people might believe. The multitudes of documented errors in both functionality and AI's ability to draw conclusions is a startling prospect when such systems are used as definitive asserters of truth in court.

B. The History of Algorithmic and AI-Based Evidence in Court

1. Risk Assessment Algorithms and Machine Learning Evidence

Machine learning takes many forms in the court system, from risk assessment and facial recognition to fingerprint analysis and generative AI. When offered as evidence, its output is generally found to be admissible under the Federal Rules of Evidence (FRE), though that admissibility is sometimes dependent upon how the algorithm was created.[52]

While predictive algorithms have become commonplace in both professional and private settings, particularly recognizable in targeted advertising, it is their usage in criminal justice which has garnered increasing controversy, especially when it concerns risk assessment.[53] The most prevalent application of this technology involves predicting recidivism rates—the likelihood that someone convicted of a crime will someday reoffend.[54] Of the various programs designed to make such predictions, COMPAS "has been used to assess over one million individuals in the criminal justice system since it was developed in 1998," and its Recidivism Risk Scale "has been in use since 2000."[55]

Reportedly Struggling to Improve Latest Models, MACRUMORS (Nov. 13, 2024, 5:30 AM), https://www.macrumors.com/2024/11/13/ai-companies-struggle-improve-llms/ [https://perma.cc/ZU68-B5UV] ("Leading artificial intelligence companies . . . are facing 'diminishing returns' . . . . Silicon Valley's belief that more computing power, data, and larger models will inevitably lead to better performance . . . could be based on false assumptions.").

52 *See* Patrick W. Nutter, Comment, *Machine Learning Evidence: Admissibility and Weight*, 21 U. PA. J. CONST. L. 919, 932 (2019) ("Machine learning output is likely admissible . . . . [T]he exact manner in which the algorithm was created or the way it would be used at trial may, in some cases, render it inadmissible.").

53 *See* Dressel & Farid, *supra* note 4, at 2–3.

54 *Id.* at 3.

55 *Id.*

To form a prediction, "COMPAS relies upon two types of data: (1) data gathered from an offenders' [sic] official record by a criminal justice professional, and (2) offenders' responses to questions that may be administered via either a paper and pencil survey or interview with a professional."[56] The individual's criminal history, employment status, age, gender, ethnicity, history of substance abuse, community ties, and level of education are all taken into account.[57] The proprietary software then uses this input data and provides an estimation of the offender's risk of violence, recidivism, and non-compliance to compute an overall "risk" score reported as either low, medium, or high as compared to other offenders.[58] For years, this information has been used during pretrial proceedings and by judges to help inform decisions on sentencing.[59]

In *State v. Loomis*, the defendant, who was involved in a drive-by shooting, received a COMPAS score indicating he presented a high risk of recidivism.[60] The State referenced this score during oral arguments as a factor that should be considered for sentencing, and the court evidently agreed.[61] "You're identified, through the COMPAS assessment, as an individual who is at high risk to the community," the judge stated.[62] "I'm ruling out probation because of the seriousness of the crime and because your history . . . and the risk assessment tools that have been utilized, suggest that you're extremely high risk to re-offend."[63]

Cases like *Loomis* are not rare. In *Santos v. Macauley*, the Sixth Circuit Court of Appeals reviewed the decision of a lower court which had considered the defendant's COMPAS risk assessment score as a factor in determining sentencing.[64] In

---

56 JENNIFER L. SKEEM & JENNIFER ENO LOUDEN, ASSESSMENT OF EVIDENCE ON THE QUALITY OF THE CORRECTIONAL OFFENDER MANAGEMENT PROFILING FOR ALTERNATIVE SANCTIONS (COMPAS) 8 (2007), https://cpb-us-e2.wpmucdn.com/sites.uci.edu/dist/0/1149/files/2013/06/CDCR-Skeem-EnoLouden-COMPASeval-SECONDREVISION-final-Dec-28-07.pdf [https://perma.cc/B792-QJDR].

57 *Justice Served? Discrimination in Algorithmic Risk Assessment*, RSCH. OUTREACH (Sept. 19, 2019), https://researchoutreach.org/articles/justice-served-discrimination-in-algorithmic-risk-assessment/ [https://perma.cc/HDQ4-WP9Z].

58 SKEEM & ENO LOUDEN, *supra* note 56, at 8, 18.

59 *See* Dressel & Farid, *supra* note 4, at 2–3.

60 State v. Loomis, 881 N.W.2d 749, 754–55 (Wis. 2016).

61 *See id.*

62 *Id.* at 755.

63 *Id.*

64 *See* Santos v. Macauley, No. 21-1076, 2021 U.S. App. LEXIS 22888, at *2 (6th Cir. Aug. 2, 2021).

denying the defendant's petition, the court explained that reasonable jurists would agree the sentencing was appropriate and the use of COMPAS was not unconstitutional.[65] The Appellate Division of New York's Supreme Court reviewed a decision by the state's Parole Board after it denied a prisoner's request to be released, citing various factors which included the prisoner's COMPAS score.[66] The court affirmed the Board's denial of parole, referencing the prisoner's predicted likelihood to return to substance abuse and his inability to prove that the Board prejudiced his rights.[67] In another, more recent case, the Court of Appeals for the Second Circuit held a petitioner's claim that his COMPAS report contained two prejudicial falsehoods was without merit, going through each contested data point to show the algorithm's assessment was not erroneous.[68] While not every court states its support of the program quite as explicitly as *Loomis*, case law makes it clear that judges are more likely to take the risk score into consideration for sentencing and parole than to question its reliability.

Facial recognition is another commonly used metric for evaluating a defendant's innocence or guilt. Described by Congress as a biometric surveillance system, facial recognition software takes multiple factors into account, including age, cosmetics, whether the individual underwent plastic surgery, the individual's pose, and the potential effects of substance abuse.[69] There are multiple programs available, each using proprietary software to match a subject's face with a database of millions, if not billions, of images from public sources and driver's license photos or mugshots.[70] A 2016 study published by the Georgetown Law Center on Privacy and Technology found that "[o]ne in two American adults is in a law enforcement face recognition network."[71] A match using facial recognition software is useful for police and criminal investigators, but the results themselves

---

65 *See id.* at *4.

66 Cassidy v. N.Y. State Bd. of Parole, 35 N.Y.S.3d 132, 134 (App. Div. 2016).

67 *Id.*

68 Amaker v. Schiraldi, 812 F. App'x 21, 24–25 (2d Cir. 2020).

69 Karissa Key, *Pros and Cons of Facial Recognition Used in Criminal Cases*, PUMPHREY L. (Sept. 4, 2023), https://www.pumphreylawfirm.com/blog/pros-and-cons-of-facial-recognition-used-in-criminal-cases/ [https://perma.cc/6R7W-J5AP].

70 *Id.*

71 Clare Garvie, Alvaro Bedoya & Jonathan Frankle, *The Perpetual Line-Up: Unregulated Police Face Recognition in America*, GEO. L. CTR. PRIV. & TECH. (Oct. 18, 2018), https://www.perpetuallineup.org [https://perma.cc/YNA5-BDBZ].

are not generally admissible as concrete evidence in court.[72] However, trial testimony regarding the use of facial recognition is permitted, and requests by defendants for discovery into the software that was used to help identify them are often denied.[73]

Finally, the rapid development of generative AI programs (GenAI), which are built upon text-generative LLMs,[74] means the possibility of litigation surrounding GenAI has become inevitable, automatically raising questions regarding the admissibility of such evidence.[75] Currently, GenAI has been involved in lawsuits involving privacy, tort, trademark, right of publicity, copyright, and facial recognition.[76] While criminal cases involving GenAI are not currently among these litigation trends, criminal justice is certainly not immune to the concerns of growing GenAI or the possibility of pleadings being written based on computer input.[77] Furthermore, the underlying concerns of ML software producing inaccurate results apply to each form of algorithmic-driven evidence and must be addressed so judges can make better-informed decisions in the future.

2. Probabilistic Genotyping

Probabilistic genotyping has emerged as a new tool employed by prosecutors, rising to prominence in part due to the "CSI Effect," which refers to a jury's greater tendency to find defendants guilty when DNA evidence is produced that ties them to the alleged crime.[78] Because probabilistic genotyping can supposedly isolate a single suspect's DNA among a sample containing DNA from multiple individuals, it can be used as

---

72 *See* People v. Reyes, 133 N.Y.S.3d 433, 436–37 (Sup. Ct. 2020) ("[A] facial recognition 'match' has never been admitted at a New York criminal trial as evidence that an unknown person in one photo is the known person in another.").

73 *See id.* at 435.

74 Elizabeth Bell, *Generative AI vs. Large Language Models (LLMs): What's the Difference?*, APPIAN (Sept. 19, 2024), https://appian.com/blog/acp/process-automation/generative-ai-vs-large-language-models.html [https://perma.cc/V2HN-C2VP].

75 *See* Maura R. Grossman et al., *The GPTJudge: Justice in a Generative AI World*, 23 DUKE L. & TECH. REV. 1, 1–2, 4 (2023).

76 Christopher J. Valente et al., *Recent Trends in Generative Artificial Intelligence Litigation in the United States*, K&L GATES (Sept. 5, 2023), https://www.klgates.com/Recent-Trends-in-Generative-Artificial-Intelligence-Litigation-in-the-United-States-9-5-2023 [https://perma.cc/7YSR-DATN].

77 *Id.*

78 *See* Daniel P. Mooney, *The Rise of Probabilistic Genotyping Causing the Fall of DNA Evidence*, MSBA (Sept. 21, 2022), https://www.msba.org/site/site/content/News-and-Publications/News/General-News/The-Rise-of-Probabilistic-Genotyping-Causing-the-Fall-of-DNA-Evidence.aspx [https://perma.cc/ZK8R-6RPN].

evidence in a greater number of cases where uncontaminated samples are unavailable.[79] However, this new approach to genotyping differs greatly from previous methods of DNA analysis.

Traditional forensic DNA analysis typically involves either loci (physical locations on a chromosome) that contain Variable Numbers of Tandem Repeats (VNTRs), or polymerase chain reaction (PCR) based analyses.[80] VNTRs are regions of DNA with large quantities of alleles—alternative versions of a gene—and, as a result, are "particularly convenient as markers for human identification" because of the high level of variation between any two individuals.[81] By extracting DNA from a sample, running it through an electrified gel assay, and comparing the fragment lengths of VNTRs, scientists can determine whether two DNA samples are a match.[82] If there is not enough of an initial sample for this method, PCR is used to "greatly amplify[] a short segment of DNA" so that alleles can be identified and compared.[83] "PCR-based methods permit the analysis of extremely tiny amounts of DNA," but that sample must be from a single individual to prevent the risk of contamination and false results.[84] These traditional methods of DNA fingerprinting are highly reliable when analyzing evidence containing a DNA sample from one person, but it becomes far more complicated when mixed samples are at issue.[85]

Technological improvements have reduced the sample size needed for analysis, but because of this heightened sensitivity, DNA from multiple individuals is often detected.[86] Ordinary DNA analysis would be unable to accurately separate each genetic profile, but probabilistic genotyping "takes incomplete or otherwise inscrutable DNA left behind at a crime scene, often in minuscule amounts, and runs it through a software program that calculates how likely it is to have come from a particular

---

79 *See id.*

80 *See* NAT'L RES. COUNCIL, THE EVALUATION OF FORENSIC DNA EVIDENCE 1, 4, 21, 216 (1996).

81 *Id.* at 14–15.

82 *See id.* at 15–17.

83 *Id.* at 21.

84 *Id.* at 23.

85 *See NIST Publishes Review of DNA Mixture Interpretation Methods*, NIST (June 9, 2021), https://www.nist.gov/news-events/news/2021/06/nist-publishes-review-dna-mixture-interpretation-methods [https://perma.cc/G26V-MZAQ].

86 JOHN M. BUTLER ET AL., DNA MIXTURE INTERPRETATION: A NIST SCIENTIFIC FOUNDATION REVIEW 11–12 (2024), https://nvlpubs.nist.gov/nistpubs/ir/2024/NIST.IR.8351.pdf [https://perma.cc/GE6M-WZLZ].

person."[87] This calculation is not perfect. According to the National Institute of Standards and Technology, while laboratories generally come to the same result when analyzing "high-quality, single-source samples," multiple interlaboratory studies over the last twenty years reveal "a wide range of results when interpreting the same *DNA mixtures*."[88]

At the forefront of this new genotyping technology is TrueAllele—a proprietary software created and sold by Cybergenetics.[89] Starting with a mixed DNA sample, the software "propose[s] tens of thousands of possible individual DNA profiles . . . [and provides] a 'likelihood ratio' that expresses the chance the suspect's DNA is in the evidence sample, relative to a random person in the population."[90] The likelihood ratio is not a conclusive match, but it provides compelling evidence and has been used in criminal trials since 2009[91] with forty crime labs approving it for use without reviewing its source code.[92] As

---

87 Lauren Kirchner, *Powerful DNA Software Used in Hundreds of Criminal Cases Faces New Scrutiny*, THE MARKUP, https://themarkup.org/news/2021/03/09/powerful-dna-software-used-in-hundreds-of-criminal-cases-faces-new-scrutiny [https://perma.cc/D64H-A4TS] (Mar. 9, 2021, 9:59 AM).

88 BUTLER ET AL., *supra* note 86, at 12.

> Distinguishing one person's DNA from another's in these mixtures, estimating how many individuals contributed to the recovered DNA sample, not knowing whether the DNA is associated with a crime or is from contamination, or whether the findings support the presence of a trace amount of suspect or victim DNA make DNA mixtures inherently more challenging to interpret than single-source samples. These issues, if not properly considered and communicated, can lead to misunderstanding the strength and relevance of the DNA evidence in a case.

Id. at 21.

89 *See* Jouvenal, *supra* note 3.

90 *Id.*

91 In 2009, TrueAllele evidence was admitted for the first time, leading to a conviction of first-degree murder for the defendant who then appealed on the basis that testimony regarding TrueAllele should have been excluded because:

> (1) "as of the date of the pre-trial hearing, no forensic laboratory in the United States used Perlin's TrueAllel [sic] method in analyzing a mixed sample of DNA for forensic purposes"; (2) "the TrueAllel [sic] system had never been used in a court of law in any jurisdiction in the United States on a mixed DNA sample to give a likelihood ratio"; and (3) no outside scientist can replicate or validate Dr. Perlin's methodology because his computer software is proprietary.

Commonwealth v. Foley, 38 A.3d 882, 888–89 (Pa. Super. Ct. 2012) (alteration in original) (citation omitted). Affirming the admission of TrueAllele evidence, the court found there was no "legitimate dispute regarding the reliability" of the evidence, novelty of a scientific method is not based on its prior usage in court, and "scientists can validate the reliability of a computerized process even if the 'source code' underlying that process is not available to the public." *Id.* at 889.

92 Jouvenal, *supra* note 3.

of this year, judges have ruled TrueAllele evidence to be admissible in over fifty cases after it was challenged by defendants at both the state and federal level.[93] As TrueAllele evidence increasingly arises in court, judges need to be prepared to make just and fair determinations about its reliability, which is only possible through an in-depth examination of the software.

III. PROBLEMS

A. Lost Transparency: Black Box Algorithms and Biased Reports

It is said that justice is blind, but in a very real sense, judges should not be. When a person's liberty hinges on evidence produced by a computer, understanding the process between the initial input data and its final output is vital to protecting justice. Yet this process is often shrouded in secrecy and hidden behind impenetrable walls in an all-too-common practice known as "black box" algorithms.[94] A program's source code dictates each action the program takes, revealing how and why it came to a certain conclusion. A lack of transparency in this regard is equivalent to an expert witness asking the court to simply take their word on something without any further explanation. And when validation studies of a program's effectiveness are offered in lieu of this explanation, they are often tainted by implicit bias. As a result, attorneys, criminal defendants, and even judges are starting to become more vocal in their protests against black box evidence, and systems like TrueAllele and COMPAS are leading offenders.[95]

These protests have arisen out of the backdrop of admissibility rules that currently govern scientific and computer-driven evidence—rules which some argue are ineffective at regulating AI. Generally, when scientific evidence is offered in trial, it is accompanied by an expert witness to explain their findings to the jury. The judge's decision whether to admit such evidence is guided by the FRE and standards explicated in two seminal cases: *Daubert v. Merrell Dow*

93 *See TrueAllele Admissibility*, CYBERGENETICS, https://www.cybgen.com/information/admissibility/page.shtml [https://perma.cc/U6PT-C9WV] (last visited Apr. 9, 2025).

94 *See* Christina Swarns, *When Artificial Intelligence Gets It Wrong*, INNOCENCE PROJECT (Sept. 19, 2023), https://innocenceproject.org/when-artificial-intelligence-gets-it-wrong/ [https://perma.cc/XNQ7-ZHBG].

95 *See* Jouvenal, *supra* note 3.

Pharmaceuticals, Inc. and *Frye v. United States*.[96] Decided in 1923, *Frye* is a District of Columbia Circuit case in which the court held that admissibility is guided by whether the principle or method from which "the deduction is made . . . [is] established to have gained general acceptance" in the relevant scientific community.[97] Seventy years later, this rule was superseded by the Supreme Court's decision in *Daubert* which shifted general acceptance in the scientific community from the only test to one of several factors judges should consider, including whether the technique can be tested for reliability, its error rate, and whether it was subject to peer review.[98] Currently, federal courts exclusively follow the *Daubert* rule while state courts remain split between the two.[99] Under either standard, black box evidence is not per se invalid.[100] While these rules may seem to be a logical foundation of admissibility in a bubble, when applied to real cases involving defendants being convicted by algorithms, cracks begin to emerge.

After a 2014 robbery at a Virginia gas station, investigators were stumped for a lead when DNA analysis on the victim's shirt resulted in zero hits.[101] Four years later, the shirt was re-tested, and improved analytical techniques revealed residual DNA of at least three individuals, but only one-third of the necessary genetic markers were present to determine a match with any one person.[102] Under traditional analysis, the evidence was a dead end; using TrueAllele, a defendant was identified and charged. There was no other evidence directly linking the defendant to the crime, and he always maintained his innocence, claiming he had never stepped foot in the county where the robbery was

96 Anjelica Cappellino, Daubert *vs.* Frye*: Navigating the Standards of Admissibility for Expert Testimony*, EXPERT INST., https://www.expertinstitute.com-resources/insights/daubert-vs-frye-navigating-the-standards-of-admissibility-for-expert-testimony/ [https://perma.cc/Z9TM-ZPCE] (Apr. 11, 2022).

97 Frye v. United States, 293 F. 1013, 1014 (D.C. Cir. 1923).

98 Daubert v. Merrell Dow Pharms., Inc., 509 U.S. 579, 592–94 (1993). Specifically, *Daubert* instructs that when "[f]aced with a proffer of expert scientific testimony, . . . the trial judge must . . . . [make] a preliminary assessment of whether the reasoning or methodology underlying the testimony is scientifically valid." *Id.* at 592–93. Pertinent considerations include: (1) whether the relevant theory or technique can be (and has been) tested; (2) whether it has been subjected to peer review and publication; (3) the theory or technique's known or potential error rate; (4) the existence and maintenance of standards controlling its operation; and (5) whether the theory or technique has attracted widespread acceptance within the relevant scientific community. *Id.* at 593–94.

99 Cappellino, *supra* note 96.

100 *See* Nutter, *supra* note 52, at 949.

101 *See* Jouvenal, *supra* note 3.

102 *See id.*

committed.[103] The public defender assigned to his case argued "it would be impossible to assess whether TrueAllele had correctly identified [the defendant] . . . without the program's source code."[104] When faced with reports questioning the program's reliability, the attorney went on to say: "We shouldn't be using the criminal justice system as a proving ground for new technologies, especially when the makers of these technologies are keeping how they work secret."[105] The defendant was indicted for robbery and use of a firearm in 2019, and in 2021, the Virginia Court of Appeals reversed the grant of bail, concluding the lower court had erred in finding he was not a danger to the community.[106] However, concerns over the use of TrueAllele were not addressed.[107] His attorney was not the first nor the last to protest the use of TrueAllele and the pervasive secrecy surrounding the software.

In *People v. Wakefield,* the defendant was convicted of first-degree murder and first-degree robbery after his DNA was identified on several items taken from the crime scene.[108] Of the evidence analyzed, lab technicians discovered a complex mixture of multiple DNA samples from which the "defendant could not be excluded," but they were ultimately unable to determine whose DNA was present with any certainty.[109] Then the data was sent to Cybergenetics.[110]

> TrueAllele concluded that it was 5.88 billion times more probable that defendant was a contributor to the mixture on the amplifier cord than an unrelated black person, . . . 170 quintillion times more probable . . . [regarding] the outside rear shirt collar, . . . and 303 billion times more probable . . . [regarding] the mixture on the outside front shirt collar.[111]

With those numbers being presented as definitive, a jury would likely have no trouble finding the defendant guilty.

Before *Wakefield* went to trial, a *Frye* hearing was held to determine admissibility of the TrueAllele evidence, and defense

---

103 *See id.*

104 *Id.*

105 *Id.*

106 Commonwealth v. Watson, No. 1284-20-4, 2021 WL 2324262, at *1–2, *5–7 (Va. Ct. App. June 8, 2021).

107 *See id.* at *2, *5–7.

108 People v. Wakefield, 195 N.E.3d 19, 21–22, 26 (N.Y. 2022).

109 *Id.* at 21–22.

110 *Id.* at 22.

111 *Id.*

counsel cited statements from Dr. Ranajit Chakraborty, a scientist specializing in evaluating methods of forensic DNA analysis.[112] Dr. Chakraborty explained TrueAllele is a "novel innovation" that has not gained "general acceptance in the scientific community," and although the program was approved by the New York State Commission on Forensic Science DNA Subcommittee, the committee had not been given proof of its analysis of complex DNA mixtures.[113] Overall, the program "ha[d] not been adequately validated for the type of casework [to which] it [was then] being applied, [and] . . . in the absence of disclosure of the source code . . . and the underlying assumptions programmed into the system, 'TrueAllele cannot be meaningfully validated.'"[114] In return, the People called several witnesses to advocate for TrueAllele's reliability, including its creator Dr. Mark Perlin, who argued that the software has been the subject of many peer-reviewed papers and underwent twenty-five validation studies.[115] Specifically calling out the software's black box nature, the defense cross-examined those witnesses, revealing that "laboratory analysts lack a complete understanding of how the . . . system works . . . [and therefore] would be [un]able to testify in court."[116] The court ultimately found the evidence to be admissible, determining TrueAllele is generally accepted in the scientific community.[117]

The studies referenced by Dr. Perlin in *Wakefield* are available on Cybergenetics' website, and a quick glance reveals one striking similarity between them—Dr. Perlin was a co-author in twenty-one out of the twenty-three listed journal

---

112 *Id.* at 23; *see also* Frye v. United States, 293 F. 1013, 1014 (D.C. Cir. 1923) (holding that the test for admissibility of scientific evidence is whether the method at issue has "gained general acceptance in the particular field in which it belongs").

113 *Wakefield*, 195 N.E.3d at 23.

114 *Id.* (first and second alterations in original). A 2024 National Institutes of Standards and Technology study had similar concerns, explaining that while there have been validation studies on probabilistic genotyping since 2014, the information found in these publications is often lacking "specific details about the samples, including the assigned [likelihood ratio (LR)] values . . . [such that] reasons for differences [among mixed DNA sample analyses] cannot be independently assessed." BUTLER ET AL., *supra* note 86, at 82, 90. A further problem is the fact that "reliability" is a more subjective than objective inquiry when it comes to LRs, and "an assessment of reliability . . . for global forensic cases [is] not feasible for LR values assigned by [probabilistic genotyping] systems . . . in large part because there is no true LR." *Id.* at 15.

115 *Wakefield*, 195 N.E.3d at 23–24.

116 *Id.* at 25.

117 *Id.*

publications.[118] The hallmark of a reliable peer-reviewed study is that it offers an independent, unbiased assessment of a particular scientific process, ensuring data was not altered or made in error by the original authors to create an artificially better result.[119] Further compounding the issue is the fact no laboratory assessing TrueAllele was given access to the program's source code, and scientists involved in determining its accuracy have admitted their lack of understanding as to how the program functions.[120] If the algorithms at issue were hidden from the very experts on which courts rely to provide an opinion, then any conclusions pertaining to the software's acceptance cannot be trusted as wholly accurate.

When arguments about general acceptance in the scientific community and due process fail, defense teams have attempted to exclude black box evidence on the grounds it violates a defendant's Sixth Amendment right to confront the witnesses against him—otherwise known as the Confrontation Clause.[121] After a 2013 double homicide in Pennsylvania, a bandana discovered at the crime scene was sent to Cybergenetics to determine if the defendant's DNA was among the mixed sample

---

118 *Publications*, CYBERGENETICS, https://www.cybgen.com/information/admissibility/page.shtml [https://perma.cc/UAP2-4WXM] (last visited Apr. 9, 2025).

119 *See Responsibilities in the Submission and Peer-Review Process,* INT'L COMM. OF MED. J. EDS., https://www.icmje.org/recommendations/browse/roles-and-responsibilities-responsibilities-in-the-submission-and-peer-peview-process.html [https://perma.cc/3SY9-2CDJ] (last visited May 4, 2025).

120 *See* Jouvenal, *supra* note 3.

> [A] judge at a previous trial . . . asked a scientist trained on TrueAllele if she could independently reproduce the results of the program . . . [and she] replied: "It would take me years to try, and I don't know that I could do it." [The scientist] went on to testify that she wouldn't be able to detect low-level errors in TrueAllele's analysis either.

Id.

121 *See, e.g.*, *Wakefield*, 195 N.E.3d at 26.

> Defendant . . . assert[ed] that the TrueAllele Casework System was the witness and that he needed the source code to effectively cross-examine that witness. . . . The court denied the request, stating that the issue defense counsel raised was a discovery issue and that defendant's ability to cross-examine Dr. Perlin . . . satisfied his right to confrontation.

Id.; Commonwealth v. Knight, No. 379, 2017 WL 5951725, at *6 (Pa. Super. Ct. Nov. 29, 2017) (finding the trial court properly denied discovery of TrueAllele's source code because it was not material to determining the program's reliability and defendant's right to confrontation was satisfied via cross-examination of Dr. Perlin); People v. H.K., 130 N.Y.S.3d 890, 897 (N.Y. Crim. Ct. 2020) (distinguishing STRMix from TrueAllele because Dr. Perlin describes TrueAllele as an "expert system" with "a certain degree of artificial intelligence," and therefore any Confrontation Clause concerns over STRMix were met by cross-examining the DNA analyst in a way that may not be satisfied with TrueAllele).

present on the cloth.[122] TrueAllele reported it was 5.7 billion times more likely the DNA belonged to the defendant than any other person, and the evidence was admitted at trial.[123] Arguing the evidence should be inadmissible without access to the source code, defense attorney Ken Haber intimated the Confrontation Clause, stating: "You can't cross-examine a computer. The Constitution demands, and justice requires, we be permitted to find out what the computer is doing to come up with its answer."[124] Haber's co-counsel, Noah Geary, had similar complaints, arguing Cybergenetics' refusal to release the source code was an "anathema to due process of law."[125] But when the defense submitted a discovery request for TrueAllele's source code, it was denied by the Court, which determined that the code is the "intellectual property of Cybergenetics," that it was neither material to the case nor necessary to evaluate the program's reliability, and that its release would cause irreparable harm to the company.[126]

TrueAllele does not stand alone in the field of secretive black box programs. Like Cybergenetics, Northpointe—the maker of COMPAS—considers the software's source code to be a trade secret and refuses to release the underlying algorithms for inspection.[127] Because the final recidivism score does not allow a user to understand how COMPAS reached that result, nor is a user able to confirm the program's supposed accuracy without its source code, "[t]he COMPAS system is not interpretable."[128] Defense teams have attempted to fight back against this secrecy in court, but without much success.

---

122 *See Trials: Commonwealth of Pennsylvania v Michael Robinson*, CYBERGENETICS, https://www.cybgen.com/news/cases/Pennsylvania-v-Michael-Robinson.shtml [https://perma.cc/9434-Y5DP] (last visited Apr. 3, 2025).

123 Paula Reed Ward, *Legal Question: How Do You Cross-Examine a Computer?*, PITT. POST-GAZETTE (Aug. 29, 2016), https://www.post-gazette.com/news/science/2016/08/29/Legal-question-how-do-you-cross-examine-a-computer/stories/201608280021 [https://perma.cc/E77Z-PT3J].

124 *Id.*

125 *Id.*

126 Memorandum Order at 1–3, Commonwealth v. Robinson, No. CC 201307777 (Pa. Ct. C.P. Feb. 4, 2016).

127 *See* Rick Jones, *The Siren Song of Objectivity: Risk Assessment Tools and Racial Disparity*, NACDL MEDIUM (July 26, 2018), https://nacdl.medium.com/from-the-president-the-siren-song-of-objectivity-risk-assessment-tools-and-racial-disparity-fa5ccb0698a5 [https://perma.cc/3823-82XQ].

128 Brandon L. Garrett & Cynthia Rudin, *Interpretable Algorithm Forensics*, 120 PROCEEDINGS NAT'L ACAD. SCIS., Oct. 2, 2023, at 1, 6.

A notable example of this strategy is found in *Loomis* wherein the defendant argued the proprietary nature of COMPAS violated his "due process right to be sentenced based on accurate information" because the validity of the COMPAS score cannot be judged without disclosure of "how the risk scores are determined or how the factors are weighed."[129] Finding his arguments unpersuasive, the court pointed to Northpointe's COMPAS manual which gives a broad overview of how the program functions and explained Loomis could still challenge the resulting score, thereby protecting due process.[130] The court then cited various validation studies purporting to show that COMPAS "is a sufficiently accurate risk assessment tool."[131] Yet, like the issues plaguing TrueAllele's validation studies, a review of risk assessment instruments in the United States revealed that "[i]n most cases, validity had only been examined in one or two studies . . . and, frequently, those investigations were completed by the same people who developed the instrument."[132]

Due process and constitutional concerns are a common refrain among attorneys and outspoken scientists[133] but often seem to fall on deaf ears as judges give deference to the business-motivated arguments of companies. The tension between protecting a defendant's right to fair justice and protecting proprietary trade secrets is only becoming more complicated as forms of AI continue to enter the courtroom at an increasingly fast rate.

B. When Machines Get It Wrong: Subjectivity and Inaccurate Results

There is a tendency to view machines as infallible arbiters of truth, particularly because, in theory, they should lack the subjective viewpoints and biases that influence human

---

129 State v. Loomis, 881 N.W.2d 749, 760–61 (Wis. 2016).

130 *See id.*

131 *Id.* at 762.

132 SARAH L. DESMARAIS & JAY P. SINGH, RISK ASSESSMENT INSTRUMENTS VALIDATED AND IMPLEMENTED IN CORRECTIONAL SETTINGS IN THE UNITED STATES 2 (2013) https://csgjusticecenter.org/wp-content/uploads/2020/02/Risk-Assessment-Instruments-Validated-and-Implemented-in-Correctional-Settings-in-the-United-States.pdf [https://perma.cc/35CF-7P7U].

133 *See* Ward, *supra* note 123 ("Dan Krane, a professor of biological sciences . . . wrote that while validation studies are important, it is the source code that serves to implement the underlying concepts of the program. 'Human experts are expected to explain how they arrive at a conclusion .this [sic] same expectation can and should apply to a computer program . . . .'").

decision-making. But AI and ML software are not without flaws, and those flaws manifest as faulty and incorrect output. Even a one percent error rate in a program's source code can "correspond to tens of thousands of errors in a single program."[134] When that error rate is applied to data being used to convict, the implications become startling, especially when keeping source codes hidden could result in errors going undetected. As explained by defense attorney Noah Geary, "[s]omething may be scientifically reliable, but that does not mean it is without flaws . . . [which] may rise to the level of reasonable doubt."[135]

A clear example of those flaws arose surrounding one of TrueAllele's largest market rivals—STRmix.[136] As a probabilistic genotyping software, STRmix claims to solve the same complicated DNA puzzles as TrueAllele, but it has undergone harsh scrutiny pertaining to its use in the Australian justice system.[137] After STRmix released its source code for inspection, Queensland authorities reported that coding errors were discovered which "affected DNA likelihood ratios in 60 cases" and prompted the replacement of STRmix evidence in twenty-four cases.[138] Though the makers of STRmix stress that each identified miscode was minor and inconsequential,[139] what may seem minor to a business focused on sales differs greatly from what a jury may view as minor in determining someone's guilt or innocence. A similar situation occurred in a U.S. criminal case when DNA evidence extracted from under the victim's fingernail was sent to both TrueAllele and STRmix in hopes of identifying the perpetrator.[140] When each software came to a different

134 Nutter, *supra* note 52, at 940.

135 Ward, *supra* note 123.

136 *See* Kwong, *supra* note 8, at 292.

137 *Id.* at 292–93.

138 David Murray, *Queensland Authorities Confirm 'Miscode' Affects DNA Evidence in Criminal Cases*, THE COURIER MAIL (Mar. 20, 2015, 10:00 PM), https://www.couriermail.com.au/news/queensland/queensland-authorities-confirm-miscode-affects-dna-evidence-in-criminal-cases/news-story/833c580d3f1c59039efd1a2ef55af92b [https://perma.cc/C8X4-FV2E].

139 *See Summary of Miscodes*, STRMIX (May 23, 2018, 9:00 AM), https://strmix.com/news/summary-of-miscodes/ [https://perma.cc/TUH6-9JR6].

140 *See* Douglass Dowty, *Judge Tosses Key Cutting-Edge DNA Before Potsdam Trial in 12-Year-Old Boy's Murder*, SYRACUSE (Aug. 29, 2016, 6:27 PM), https://www.syracuse.com/crime/2016/08/judge_tosses_cutting-edge_dna-before_potsdam-trial_in_12-year-old_boys_murder.html [https://perma.cc/M7JZ-RC7V].

conclusion after analyzing the same DNA sample, the judge ruled the evidence was inadmissible.[141]

If each program was as accurate as their marketing materials proclaim, they should reach the same conclusions when analyzing the same DNA since their respective companies purport to use equivalent methods of probabilistic genotyping. Yet inconsistencies between results is not uncommon. A study conducted by the Department of Criminology, Law and Society at the University of California, Irvine (UCI) used TrueAllele and STRmix to analyze a mixed DNA sample that traditional methods of analysis could not match.[142] Each program produced startling different results. TrueAllele presented four values comparing the DNA to various racial groups with a likelihood of a match between the sample and the defendant ranging from 1.2 million to 6.07 million times less probable than a "coincidental match" to another person.[143] STRmix determined the sample was twenty-four times more likely to have originated from two unknown contributors other than the defendant.[144] While both concluded the defendant's DNA was likely not in the mixture, TrueAllele's likelihood ratio was "larger by five to six orders of magnitude."[145] One's guilt or innocence should not hinge on which program was used to analyze evidence.

The author of the UCI study attempted to account for the discrepancy, pointing to variations in how each program sets certain values for analysis, as well as "misleading" ways in which the results are presented.[146] Specifically calling out TrueAllele, he explained, "[b]ecause Cybergenetics is using the terms 'match' and 'coincidence' to convey a meaning that is very different from what most people think those terms mean, and because Cybergenetics fails to explain this departure . . . I believe that Cybergenetics' LR [likelihood ratio] statement is inappropriate and misleading."[147] Overall, the study highlighted significant

---

141 Lauren Kirchner, *Where Traditional DNA Testing Fails, Algorithms Take Over*, PROPUBLICA (Nov. 4, 2016, 8:00 AM), https://www.propublica.org/article/where-traditional-dna-testing-fails-algorithms-take-over [https://perma.cc/N4Q7-FKQ4].

142 William C. Thompson, *Uncertainty in Probabilistic Genotyping of Low Template DNA: A Case Study Comparing STRMix and TrueAllele*, 68 J. FORENSIC SCIS. 1049, 1051 (2023).

143 *Id.* at 1053.

144 *Id.*

145 *Id.* at 1054.

146 *Id.*

147 *Id.* at 1059. "Misleading" seems to be an appropriate term for the LRs expressed in probabilistic genotyping software. While most lay readers and lawyers might assume it

problems in both how these programs are implemented and how they are interpreted by lawyers and jurors, warning that statistical models resting on "unrealistic assumptions" will produce false results.[148]

In a paper published by the same author only a few months later, Dr. Thompson responded to criticism of his study, specifically calling out Dr. Perlin for making misleading statements regarding TrueAllele's reliability and for insinuating there were errors in his conclusions.[149] Explaining that his original article "raised a number of additional concerns about [TrueAllele] that Dr. Perlin and his colleagues failed to address,"[150] Dr. Thompson questioned Dr. Perlin's claim that TrueAllele is a fully automatic, Bayesian-statistics-based[151] system with "no need for analytic thresholds."[152] Unable to agree with Dr. Perlin's assessment that TrueAllele would always produce reliable, trustworthy data, Dr. Thompson invoked a sentiment common among computer scientists: "garbage in-garbage out."[153] It is unclear "at what point[] the

represents the likelihood of a proposition being true—of the defendant's DNA being present in the mixed sample—that is not the case. Instead, it represents the "ratio of the probability of the findings given [hypothesis one] is true," namely that the defendant's DNA contributed to the sample, "versus the probability of the findings given [hypothesis two] is true," that someone other than the defendant contributed to the sample. BUTLER ET AL., *supra* note 86, at 49–50. If the LR is one hundred, then the chance of seeing the exact DNA results present in the sample is one hundred times more likely to occur if the defendant's DNA is part of the sample rather than another person's. It does not mean that it is one hundred times more likely that the defendant's DNA is actually in the sample. The misinterpretation of this value is known as "transposing the conditional," or the "prosecutor's fallacy," in which a person "confuses 'the probability of the evidence given the propositions' with 'the probability of the propositions given the evidence.'" *Id.* at 50–51. For example, instead of DNA evidence, the evidence at issue is Earth's sky. And instead of the proposition being that the defendant's DNA is in the DNA evidence, your proposition is that the sky appears blue. The LR would represent the probability that you're looking at Earth's sky rather than the sky on Mars, given the proposition that the sky is blue. It does not tell you the likelihood of the sky actually being blue.

148 *See* Thompson, *supra* note 142, at 1050, 1060.

149 *See* William C. Thompson, *Author's Response*, 69 J. FORENSIC SCIS. 1519, 1519 (2024).

150 *Id.* at 1521.

151 Based on Bayes' theorem, which "is a mathematical formula that determines the conditional probability of any given event," Bayesian statistics is an approach to data analysis whereby "available knowledge regarding parameters in statistical models is updated using the information gathered from observed data." John Terra, *What Is Bayesian Statistics, and How Does It Differ from Classical Methods?*, CALTECH, https://pg-p.ctme.caltech.edu/blog/data-science/what-is-bayesian-statistics [https://perma.cc/5XKB-XT7C] (Aug. 14, 2024). This means that probabilities are continuously refined as more evidence becomes available. *See id.*

152 Thompson, *supra* note 149.

153 *Id.*

LRs produced by [TrueAllele] become garbage," and research has already suggested TrueAllele is *not* reliable for all types of DNA mixed samples.[154]

The larger problem with probabilistic genotyping like TrueAllele is the inherent subjectivity involved in certain aspects of the process—subjectivity that is now governed by computer software hidden behind claims of trade secrets. "Numerical results obtained from assigning LR values are dependent on the evidence available, statistical models applied, propositions selected based on case information, and the scientist making various judgments. . . . [which means] results vary based on [the] amount of information available and assumptions made."[155] The way in which the program is implemented directly affects the program's output and the corresponding result offered in court. In general, probabilistic genotyping systems compute LR based on:

> (1) *modeling choices* made by the system architect(s), (2) *data input choices* made by the analyst regarding an analytical threshold for calling peaks as alleles, selecting the number of contributors to the mixture for use in PGS calculations, and sometimes categorizing artifacts (e.g., pull-up peaks), (3) *proposition choices and assumptions* made by the analyst (e.g., use of unrelated individuals versus relatives, conditioning on a victim when analyzing an intimate sample, . . . and underestimating or overestimating the number of contributors), and (4) *population database choices* used by the laboratory to provide allele and genotype frequency estimates.[156]

When a forensic scientist makes those necessary choices and judgments, he or she can explain to the court which choices were made and why. When a computer running black box software does so, the court is left in the dark. In fact, the need for certain judgments to be made during the calculation process is a contributing factor as to why various probabilistic genotyping programs will come to vastly different conclusions when analyzing the same evidence.[157] And if there are flaws in the coding that

---

154 *Id.* Studies have shown direct evidence of TrueAllele producing "falsely exculpatory" LRs, finding that a defendant's DNA was more likely to not be present in a mixed sample when, in reality, it *was* included in the sample. Validation studies cited by Dr. Perlin in response to these concerns do not address this particular issue and in fact "establish the opposite"—"[t]hey show that exculpatory results of this type are often NOT accurate and hence cannot be trusted." *Id.* at 1519–20.

155 BUTLER ET AL., *supra* note 86, at 48.

156 *Id.* at 51.

157 *See id.* at 52–53.

directs the program on how to judge certain aspects of DNA evidence, there are corresponding flaws in the output—flaws that go unseen and unchecked under the current system for admitting machine evidence.

In the context of risk assessment tools, the potential for false conclusions has already been realized and proven in several studies analyzing the accuracy of their predictions. In 2016, *ProPublica* published a groundbreaking review of COMPAS, bringing the program under heavy inquiry after it was revealed to produce discriminatory and biased outputs.[158] Comparing two defendants, one African American with only juvenile misdemeanors and one Caucasian with armed robbery and attempted armed robbery convictions, COMPAS predicted the African American had a higher likelihood of recidivism.[159] Two years later, that defendant had no further charges while the Caucasian defendant was serving an eight-year prison sentence for robbery; COMPAS "got it exactly backward."[160]

Digging deeper into the program, *ProPublica* obtained the risk scores assigned to seven thousand defendants in Florida from 2013 to 2014 and found COMPAS had a success rate of only twenty percent when it came to accurately predicting which would go on to commit further crimes.[161] There were also "significant racial disparities" that remained unaffected by records of prior offenses, and "the scores ma[de] little sense even to defendants."[162] Taking *ProPublica*'s research a step further, Dr. Melissa Hamilton of the University of Surrey analyzed COMPAS's predictive validity in terms of gender, "proving that the tool overpredicts the risk for women to reoffend, therefore leading to unfair penalties for female offenders."[163] Dr. Hamilton

---

> Likelihood ratios are assigned and not measured. Different individuals may assign different LR values, even when using [probabilistic genotyping] systems, when presented with the same evidence because they base their judgments on different collection protocols, quantification systems, STR kit results, interpretation protocols, models, assumptions, or computational algorithms. For any given sample, there is no single, true likelihood ratio.

Id. at 54.

158 *See* Angwin et al., *supra* note 6.

159 *Id.*

160 *Id.*

161 *Id.*

162 *Id.*; *see also* DESMARAIS & SINGH, *supra* note 132, at 49 ("[P]erformance within and between [risk assessment] instruments varie[s] considerably depending on the assessment sample, circumstances, and recidivism outcome.").

163 *Justice Served? Discrimination in Algorithmic Risk Assessment*, *supra* note 57.

attributed this disparity in part to the fact the ML algorithm had been trained using samples of primarily male offenders, thereby limiting the scope of the program's capabilities and increasing the potential for bias.[164]

But it seems the more companies attempt to manipulate programming and data sets to increase objectivity and reduce bias, the less objective these programs become. Recently, Google launched an updated LLM called Gemini, which "produce[d] images of Black, Native American and Asian people when prompted – but refuse[d] to do the same for White people," citing racial concerns.[165] When Gemini was asked to show historical pictures of Nazi soldiers during World War II, it only produced images of people of color wearing the distinctive Nazi uniform, unable to understand the historical inaccuracy of the results.[166] Google apologized for the debacle and removed Gemini to "improve" its programming,[167] but the situation provides a clear, visual example of how coding can manifest unanticipated mistakes.

Empirical data shows there is not a single sector of ML or AI-driven software free from errors. While facial recognition has high accuracy in an ideal environment, real-world conditions are often far from perfect and result in error rates anywhere from

---

164 *See id.*

165 Nikolas Lanum, *Google Apologizes After New Gemini AI Refuses to Show Pictures, Achievements of White People*, FOX BUS. (Feb. 21, 2024, 2:32 PM), https://www.foxbusiness.com/media/google-apologizes-new-gemini-ai-refuses-show-pictures-achievements-white-people [https://perma.cc/E69B-UL5J].

166 *See* Adi Robertson, *Google Apologizes for 'Missing the Mark' After Gemini Generated Racially Diverse Nazi*, THE VERGE (Feb. 21, 2024, 2:17 PM), https://www.theverge.com/2024/2/21/24079371/google-ai-gemini-generative-inaccurate-historical [https://perma.cc/QBV9-5JYN]. Programmers have attempted to fix inaccuracies such as this with self-correction tools, but "LLMs sometimes actually perform worse with self-correction measures," and "self-correction isn't consistently effective." Samantha Keefe & Thomas Gaitley, *AI Self-Correction*, LIONBRIDGE (Jan. 16, 2024, 9:30 AM), https://www.lionbridge.com/blog/translation-localization/ai-self-correction/ [https://perma.cc/3S3T-FKA2]. While certain issues associated with LLMs like Gemini are not wholly relevant to the operation of programs like TrueAllele or COMPAS, which do not use GenAI to produce output, their inaccuracies highlight a larger problem: how can any form of machine evidence be trusted when the very programmers designed to ensure accurate output are unable to fix even a problem as obvious as non-White Nazi officers? The building blocks of Gemini, TrueAllele, and COMPAS are the same—algorithms and lines of code telling the computer what to do—but just because errors are more clearly visible in GenAI does not mean errors do not exist in programs like COMPAS; simply, those errors are more easily hidden.

167 *See* Robertson, *supra* note 166.

9.3% to 64%.[168] These programs have also been shown to misidentify female and minority populations at a disproportionate rate, often because of the skewed and incomplete information used to train the underlying algorithms.[169]

Gaining heightened media attention is the phenomenon of hallucinations plaguing the field of GenAI. While the nature of LLMs makes it difficult to determine exactly how often fabricated results are produced, estimates reveal hallucinations occur up to 27% of the time[170] and "leading AI experts aren't entirely sure what causes hallucinations."[171] In an extreme example of this issue, a New York attorney used a GenAI program to assist in writing his trial brief, and six of the referenced cases were entirely contrived "with bogus quotes and bogus internal citations."[172] When directly asked, the program assured the user that the cases were real and even provided full citations and instructions to locate them on Westlaw and LexisNexis.[173]

Hallucinations are not directly comparable to the types of problems plaguing machine evidence like TrueAllele, but the phenomenon raises an important point about the inherent trustworthiness, or lack thereof, of machine evidence, especially those with hidden source codes. It also raises an even larger issue inherent in any discussion about the use of AI in court—the implicit, misplaced trust lawyers (and judges) have in computer output. "If a computer said it, it must be true" seems to be the prevailing attitude, so much so that a lawyer relied on GenAI to provide him a trial brief without once considering that the

168 *See* William Crumpler, *How Accurate Are Facial Recognition Systems – and Why Does It Matter?*, CTR. FOR STRATEGIC & INT'L STUD. (Apr. 14, 2020), https://www.csis.org/blogs/strategic-technologies-blog/how-accurate-are-facial-recognition-systems-and-why-does-it [https://perma.cc/W5X2-N7J4].

169 *See* John McNichols, *How Do You Cross-Examine Siri if You Think She's Lying?*, AM. BAR ASS'N (May 24, 2022), https://www.americanbar.org/groups/litigation/resources/litigation-news/2022/how-do-you-cross-examine/ [https://perma.cc/N657-L28L].

170 *See* Stefan Bardega, *Generative AI Hallucinations: How Often Do They Happen and Should Marketers Be Worried?*, IDX (Nov. 8, 2023), https://www.idx.inc/blog/technology/gen-ai-hallucinations [https://perma.cc/W7QV-KZFY].

171 *Generative AI Hallucinations: Why They Occur and How to Prevent Them*, TELUS DIGIT. (July 6, 2023), https://www.telusinternational.com/insights/ai-data/article/generative-ai-hallucinations [https://perma.cc/8V2W-T7NS].

172 Ramishah Maruf, *Lawyer Apologizes for Fake Court Citations from ChatGPT*, CNN BUS. (May 28, 2023, 3:28 PM), https://www.cnn.com/2023/05/27/business/chat-gpt-avianca-mata-lawyers/index.html [https://perma.cc/5YZ4-VFZL].

173 *See id.*

information in the brief could be inaccurate.[174] Trusting faulty citations is one thing; trusting faulty conclusions about a defendant's guilt or innocence is quite another.

Computer programming has greatly advanced over the decades, but it must be remembered that algorithms are in no way perfect. "The human is in the software in the source code,"[175] and humans make mistakes. Miscodes, incomplete training data, and underlying biases all contribute to incorrect output, making AI less "intelligent" than one might assume and highlighting the danger in blindly trusting a computer in court.

IV. PREVIOUSLY PROPOSED SOLUTIONS AND WHY NONE ARE THE ANSWER

A. Revising Current Admissibility Standards

When admitting AI-driven and machine evidence, it can be judged under one of three broad standards—"direct witness testimony, expert witness testimony, or measurement using established technology"[176]—but no method on its own provides the level of rigor necessary to ensure reliability. Most scholars seem to agree machine evidence should be evaluated in the same manner as previous technological advancements, but some have proposed revising current admissibility guidelines to carve out unique rules applicable to AI. Not only would this require amending the FRE or the *Daubert* and *Frye* standards, but it would not solve many of the overarching issues outlined in Part III.

Several authors have suggested revisions to either FRE 901, which governs authentication,[177] or rules 702 to 704, which provide guidelines for admitting expert testimony and implicate the judge-made rules in both *Daubert* and *Frye*.[178] As written,

---

174 This sentiment only appears to be strengthening despite the multitude of studies demonstrating the dangers of relying on various forms of AI. Thomson Reuters, a company well-known for providing information and services within the legal profession, has published articles praising the advent of AI in the legal system, even going so far as to claim AI can "improve equity and reduce bias in judicial outcomes." Allyson Brunette, *Humanizing Justice: The Transformational Impact of AI in Courts, from Filing to Sentencing*, THOMSON REUTERS (Oct. 25, 2024), https://www.thomsonreuters.com/en-us/posts/ai-in-courts/humanizing-justice/ [https://perma.cc/UD9W-KJ92] ("Artificial intelligence (AI) tools are being introduced at every step of [the legal system] . . . . [to] improve[] efficiency and equity for defendants and their legal representation.").

175 Ward, *supra* note 123.

176 Paul W. Grimm, Maura R. Grossman & Gordon V. Cormack, *Artificial Intelligence as Evidence*, 19 NW. J. TECH. & INTELL. PROP. 9, 79 (2021).

177 FED. R. EVID. 901.

178 *See* FED. R. EVID. 702–04.

FRE 901 broadly governs how to authenticate evidence—how to prove a piece of evidence actually is what a party says it is.[179] Before the development of AI, this was a fairly simple task; now, when generated images and text are sometimes indistinguishable from the same materials produced by a human, it becomes far more complicated. FRE 702 to 704 lay out guidelines for the admissibility of expert testimony, including the expert's specialized knowledge in the subject and whether the testimony is based upon "sufficient facts or data" and "is the product of reliable principles and methods."[180] When scientific evidence was produced by a technician in a lab, this rule was a logical way of ensuring reliability. With AI, it is unclear whether any expert can truly have detailed knowledge as to how a black box program functions or testify regarding its reliability without access to its code, and an increasing number of scholars have begun to ask these pertinent questions.

Specifically addressing the problem of deepfakes, which are highly realistic AI-generated images, videos, and audio,[181] law professor Rebecca Delfino argues that the current FRE is incapable of meeting the challenges of this new category of evidence.[182] Like hallucinations, deepfakes are not a direct result of systems like TrueAllele or COMPAS, but they tend to show a clear, visual example of the pervasive problems posed by forms of AI—problems that are only exacerbated by programs purporting to definitively prove "who done it." Delfino critiques the current division of responsibility between the judge and jury for authenticating evidence, pointing to the fact jurors may be convinced by false AI evidence or allow personal skepticisms or biases to govern their decisions, and instead proposes tipping the balance in favor of judges.[183] FRE 901 "should be amended to add a new subdivision (c) . . . [that would] expand the gatekeeping function of the court by assigning the responsibility of deciding

179 FED. R. EVID. 901(a) ("To satisfy the requirement of authenticating or identifying an item of evidence, the proponent must produce evidence sufficient to support a finding that the item is what the proponent claims it is.").

180 FED. R. EVID. 702(a)–(d).

181 Ian Sample, *What Are Deepfakes – and How Can You Spot Them?*, THE GUARDIAN (Jan. 13, 2020, 5:00 PM), https://www.theguardian.com/technology/2020/jan/13/what-are-deepfakes-and-how-can-you-spot-them [https://perma.cc/266Y-9VKT].

182 *See* Rebecca A. Delfino, *Deepfakes on Trial: A Call to Expand the Trial Judge's Gatekeeping Role to Protect Legal Proceedings from Technological Fakery*, 74 HASTINGS L. J. 293, 332 (2023).

183 *See id.* at 336–37, 341.

authenticity issues solely to the judge."[184] The author argues this expanded role of the judge's fact-finding responsibilities is not without precedent, citing traditional English law which gives judges broad plenary power "to decide all questions of fact conditioning the admissibility of testimony."[185] This new division of authority to determine authenticity is meant to solve the problem of admitting realistic deepfakes and the potentially cost prohibitive nature of proving whether an image is real or AI-generated.[186]

Although Delfino articulates several poignant issues regarding FRE 901's application to AI evidence, her solution is not without flaws. For Delfino's plan to be most effective, the judge would need a requisite understanding of AI and deepfake technology in order to properly determine whether such evidence was more likely authentic or falsified. Yet the language of the suggested amendment is silent on this issue, and the article overall offers no proposed mechanism for judges to obtain that expertise.[187] This problem is exacerbated by the fact that according to Rule 2.9(c) of the Model Code of Judicial Conduct—a rule over thirty states have adopted[188]—judges are not authorized to independently research facts.[189] They therefore could not gain a meaningful understanding of deepfake technology adequate enough to singlehandedly decide issues of AI admissibility without opinions from an expert.

Taking a broader approach to address AI generally, another law professor, Victor Metallo, focuses on the rules surrounding expert scientific testimony and suggests amending the *Daubert* standard as well as the FRE's guidelines for expert testimony

---

184 *Id.* at 341.

185 *Id.* at 342.

186 *See id.* at 340.

187 *See id.* at 341.

> Notwithstanding subdivision (a), to satisfy the requirement of authenticating or identifying an item of audiovisual evidence, the proponent must produce evidence that the item is what the proponent claims it is in accordance with subdivision (b). The court must decide any question about whether the evidence is admissible.

Id. (suggesting amendment to FED. R. EVID. 901(c)).

188 KEITH R. FISHER, NAT'L CTR. FOR STATE CTS., NEW ABA ETHICS OPINION EXPLORES THE PROHIBITION ON INDEPENDENT FACT RESEARCH BY JUDGES 2 (2018), https://www.ncsc.org/__data/assets/pdf_file/0022/19309/new-aba-ethics-opinion-explores-prohibition-independent-fact-research-by-judges.pdf [https://perma.cc/G66H-LX3B].

189 MODEL CODE OF JUD. CONDUCT r. 2.9(c) (AM. BAR ASS'N 2020).

under FRE 702.[190] Metallo argues the factor-based *Daubert* test, which was designed to ensure "that any and all scientific testimony or evidence admitted is not only relevant, but reliable,"[191] is only adequate for machine evidence akin to calculators that simply speed up a human function.[192] It is, however, insufficient to tackle situations "where the machine is the unique source of knowledge, [and] where the human being might be just a vector delivering that knowledge."[193] To solve this conundrum, Metallo proposes FRE 702 should be amended to include "reliability requirements for AI" while also ensuring AI cannot wholly replace human expert testimony, only assist it.[194]

However, this proposed amendment does not fully address the root problem concerning AI evidence and could potentially create confusion or inconsistency—an outcome contrary to Metallo's goal of avoiding conflicting admissibility decisions under the current FRE.[195] Implementing "reliability requirements" could be highly beneficial, but the phrase lacks explanation as to what those requirements might be or how the courts should weigh various factors that may point towards a program's reliability. Should there be a delineated error rate over which no evidence is admissible? Would certain types of ML be afforded more protection than others? Without answering these questions, there can be no assurance such a change will effectively result in more consistent decisions.

Metallo also calls for a further amendment to the FRE which would permit the court to deem testimony inadmissible under the following circumstances:

> (1) a judge cannot take judicial notice of an Al process; or (2) where a party has not proffered an engineer to assist in explaining the Al's processes to a jury; or (3) where the Al has reached a point that "black box" processes cannot be explained by human testimony, because Al has adapted the ability to program itself.[196]

Analyzing this secondary proposal, more questions begin to emerge. Basing discretion on whether the judge can "take judicial

---

190 *See* Victor Nicholas A. Metallo, *The Impact of Artificial Intelligence on Forensic Accounting and Testimony—Congress Should Amend "The* Daubert *Rule" to Include a New Standard*, 69 EMORY L.J. ONLINE 2039, 2041–42 (2020).

191 Daubert v. Merrell Dow Pharms., Inc., 509 U.S. 579, 589 (1993).

192 Metallo, *supra* note 190, at 2048–49.

193 *Id.*

194 *Id.* at 2060.

195 *See id.*

196 *Id.* at 2061.

notice of an AI process" does not consider black box issues wherein no one can take notice of a program's process, nor does it address the judge's inability to research facts about such a process on their own. But arguably the most problematic aspect is subsection (3). As written, the amendment suggests AI may be capable of programming itself, yet most computer scientists agree that while certain AI models have been trained to mimic code written by humans, they cannot learn it to the degree necessary to create new programs.[197] "The computer appears to 'understand' things . . . [but] it wouldn't understand ANY of that if human programmers hadn't first painstakingly taught it how," one software developer explains.[198] "AI systems can't even *learn* on their own. . . . [and e]ven with new-fangled quantum hardware, I would seriously question a computer's ability to come up with innovative code . . . Behind every successful AI is a programmer rolling their eyes."[199]

The misunderstanding of the tenets of computer science implicated by this proposal demonstrates the danger of entrusting the evaluation of machine evidence solely to lawyers and judges whose areas of expertise are in a far different field. While the current rules may not address AI outright, no suggested revision appears to be without flaws that could further complicate an already convoluted area of evidence.

B. Maintaining the Status Quo

Rather than developing new admissibility rules each time there is a technological advancement that affects the legal system, others suggest keeping the rules as they are and simply adjusting their application when AI-driven evidence is at issue. The wheels of justice are known to move slowly; those of science try to break the speed limit. Because the FRE and its state counterparts are revised infrequently and via an extremely lengthy process, it is impractical to set a standard of revising the rules each time a new form of evidence is developed.[200] As a result, certain scholars argue there is nothing inherently inadequate about applying the rules as they stand. But leaving the status quo in place without developing a new method of

197 *See* Tim Baker, *What Artificial Intelligence Can't Do*, MEDIUM (Sept. 12, 2023), https://medium.com/codex/what-artificial-intelligence-cant-do-b92b4ddcf8b3 [https://perma.cc/C2M9-UBF6].

198 *Id.*

199 *Id.*

200 *See* Grossman et al., *supra* note 75, at 16.

evaluation better equipped to handle this emerging field of evidence is not without its problems as well.

The standards for expert testimony provide a clear example of this problem. The FRE allows for such testimony if the expert's "scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue,"[201] and if their opinion is based on "facts or data in the case that the expert has been made aware of or personally observed."[202] The Supreme Court has explained these rules grant expert witnesses "testimonial latitude . . . on the assumption that the expert's opinion will have a reliable basis in the knowledge and experience of his discipline."[203] While this assumption may be acceptable in other areas of expertise, it rests on shakier ground with programs that do not require user input, or even user understanding, to reach a result. In the Sixth Circuit case *U.S. v. Ganier*, the government argued against the admissibility of expert testimony based on computer search results, contending such testimony "is not based on scientific, technical, or other specialized knowledge, but is simply lay testimony available by 'running . . . software, obtaining results, and reciting them.'"[204] Using AI is no different than using a search engine in this respect—a user inputs information and receives an answer without necessarily knowing or understanding the steps in between. As written, the FRE does not address this shift from the traditional definition of expert testimony. Yet many scholars propose that the current system of deciding admissibility does not need alteration and is more than capable of handling AI.

Focusing on the admissibility of ML evidence, author Patrick Nutter surveys the existing rules before concluding "there is nothing inherently inadmissible about ML evidence under the Federal Rules of Evidence, the Fifth Amendment, or the Sixth Amendment."[205] He recognizes the potential dangers in allowing such evidence to have blanket admissibility, but rather than amending any existing standard, he suggests allowing the trier of fact to decide for themselves what weight to give computer evidence.[206] Briefly touching on the unexplainability of much of

201 FED. R. EVID. 702(a).
202 FED. R. EVID. 703.
203 Kumho Tire Co. v. Carmichael, 526 U.S. 137, 138 (1999).
204 U.S. v. Ganier, 468 F.3d 920, 925 (6th Cir. 2006).
205 Nutter, *supra* note 52, at 949.
206 *See id.* at 919.

that evidence, especially concerning black box programs, Nutter determines the "Sixth Amendment merely requires that the evidence be introduced with expert testimony."[207]

His paper provides an excellent outline of the significant evidentiary issues plaguing ML but does not seem to give an in-depth examination as to how the current setup is sufficiently designed to address them. Though the author cautions that judges and juries should be wary when weighing the strength of computer evidence, there is no explicit guidance regarding how they should do so nor whether certain types of AI require more detailed analysis than others. Overall, this position essentially leaves the situation exactly as it stands, solving none of the problems many scholars claim will soon be impacting most litigators.[208]

Taking a similar stance to Nutter, the authors of a paper, written in part by former U.S. District Court Judge Paul Grimm, argue the FRE is adequate for evaluating machine evidence without amendment "provided [the rules] are applied flexibly."[209] The authors emphasize the importance of demonstrating whether the AI evidence can be trusted as accurate before judges are then free to use their already broad discretion in deciding questions of admissibility.[210] Specifically, if a party plans to offer AI evidence, they must do so in advance of trial to ensure it meets "adequate thresholds of validity and reliability" before being presented to the jury.[211] When black box evidence is at issue, and the program's source code is not revealed, the authors explain that the party arguing for its admissibility must demonstrate validity and reliability some other way.[212]

Although reliability thresholds would be an important step forward, their paper leaves many questions unanswered and does not provide clear guidelines on how to evaluate certain types of machine evidence nor how a party would go about adequately proving a program's reliability. It places the burden entirely on

---

207 *Id.* at 958.

208 *See id.* at 919 ("Artificial intelligence ('AI') is gaining traction in legal practice. How prosecutors prioritize which crimes to prosecute, sift through mountains of documents, and establish reasonable suspicion can all reasonably be expected to change with coming AI technologies."); *see also* Delfino, *supra* note 182, at 297 (arguing that AI "will soon make trial attorneys' and judges' jobs significantly more challenging" and will require "additional measures" to evaluate).

209 Grimm, Grossman & Cormack, *supra* note 176, at 85.

210 *See id.* at 104.

211 *Id.* at 89.

212 *Id.*

the court system, yet the authors themselves admit, "When it comes to technical evidence like AI, the judge often is in a battle of wits unarmed," and leaving the obligation to attorneys "can be a challenge for lawyers who . . . [are] not specialists in the many scientific and technical disciplines that underlie AI systems."[213] They acknowledge the inherent difficulties in relying on current admissibility standards, yet ultimately argue those same standards are more than capable of handling ever-evolving AI evidence.

This position is also troublesome when it comes to excluding hearsay. FRE 801 defines hearsay as an assertion made outside the current trial, which is offered to "prove the truth of the matter asserted in the statement."[214] Because "statement" is defined as a "person's" assertion,[215] the authors distinguish AI-generated output because it is not a direct human assertion and therefore any issues of hearsay are irrelevant.[216] To support this position, the authors cite multiple federal cases that echo the sentiments of the Fourth Circuit, which determined "[o]nly a *person* may be a declarant and make a statement," and therefore, nothing "said" by a machine can qualify as hearsay.[217] As such, the authors dismiss the concept out of hand, concluding there can be no route to finding AI evidence inadmissible via hearsay.

Yet this viewpoint ignores the human behind the machine, which inherently complicates the question of whether AI can assert in a way simple calculators cannot. Scientists have written extensively on the issue of computer assertions in this new age of AI, cautioning that it is not the AI but "the humans who employ such systems who are responsible and sanctionable for the outputs and their effects."[218] Excluding AI from the protections of FRE 801 is a dangerous precedent, and the number of courts which have done just that demonstrates the need for an updated system of analyzing computer evidence. As another law professor argued, the "lack of understanding as to how" modern AI makes decisions will result in admitting unreliable statements because "the testimonial risks that are inherent in statements made by modern Al Entities are more akin to those found in human

213 *Id.* at 88–89.

214 FED. R. EVID. 801(a)–(c)(2).

215 *Id.* at 801(a).

216 *See* Grimm, Grossman & Cormack, *supra* note 176, at 85–86.

217 United States v. Washington, 498 F.3d 225, 231 (4th Cir. 2007).

218 Patrick Butlin & Emanuel Viebahn, *AI Assertion*, 2023 ERGO: OPEN ACCESS J. PHILOSOPHY, at 1, 24.

assertions that render them hearsay."[219] Providing blanket immunity from hearsay analysis to all types of machine output would allow individuals to have a computer "assert" on their behalf while being safe in the knowledge it will be admitted as objective evidence.

While amending the standards of admissibility would be a complicated and lengthy process, relying on the same rules that allow potentially faulty AI evidence to enter trials unquestioned and unimpeded is just as problematic. Ultimately, a problem cannot be solved using the same methods which created it, and AI admissibility requires a new way of thinking and a new, uniform answer.

V. A NEW SOLUTION

A. A Federal Agency and Court-Appointed Advisors

1. Filling in the Gaps of Previously Proposed Solutions

A review of the current literature on AI evidence quickly reveals there is little to no consensus regarding how to deal with this burgeoning problem, nor which organization should be responsible for doing so. If computer-driven evidence is to become the norm, there must be a recognizable standard across the court system for how to address it in order to preserve fair and consistent justice. A federal agency would do just that. Unlike potential rule revisions that fail to address every aspect of the problem or only serve to create more confusion, the solution of expert advisors organized by a centralized power would not suffer from those same drawbacks. And rather than leaving the situation alone and hoping these issues will resolve themselves over time, this provides a real mechanism for beneficial change.

Two pervasive issues plaguing each previous solution—ones that were even acknowledged by multiple authors—are the fact judges generally lack the training and experience necessary to understand computer programming, and black box programs preclude the possibility of understanding entirely. Usually, this is resolved by offering expert testimony at trial, but as discussed above, it is not as straightforward a process when it comes to AI, particularly AI operating via hidden source codes. Having a mechanism already in place to answer judges' questions about

219 Jess Hutto-Schultz, *Dicitur Ex Machina: Artificial Intelligence and the Hearsay Rule*, 27 GEO. MASON L. REV. 683, 685 (2020).

machine evidence, one which is subject to strict rules of confidentiality, solves both issues.

Rather than forcing defendants to foot the bill of expensive experts, advisors sent by the agency would already be trained and available for use, reducing costs and increasing court efficiency. And instead of each party having to comply with complicated non-disclosure agreements, assuming companies agreed to release programming materials to them at all, a mechanism would already be in place to facilitate cooperation between those companies and the court system. By subjecting the agency and its advisors to strict confidentiality, companies' concerns of proprietary interests would be assuaged, more so than if an individual expert with potentially thievish motivations was given access to source codes.[220] A centralized agency would also address the fact that expert witnesses hired by a particular side carry an inherent risk of the witness being biased for that side.[221]

This solution aims not to revise the current admissibility rules that have proven to be more than adequate in other areas of evidence, but to provide an additional safeguard on top of the existing system. As explained by one researcher: "To expect competing for-profit companies to refrain from overclaiming and to fully disclose all uncertainties surrounding their findings is apparently expecting too much. To expect courts to regulate these matters as part of their review of admissibility apparently is also expecting too much."[222] The best solution is to filter machine evidence through an extra hurdle to ensure juries are not relying on faulty or misleading evidence represented as infallible truth. Defense counsels will not be forced to take a crash course on computer science to even hope to understand the technical documents behind a program's function. And judges' inability to

220 Though individuals can seek information from federal agencies under the Freedom of Information Act, it specifically carves out an exception for "trade secrets and commercial or financial information" that is designed "to protect the interests of both the government and submitters of information." *FOIA Guide, 2004 Edition: Exemption 4*, U.S. DEP'T OF JUST., https://www.justice.gov/archives/oip/foia-guide-2004-edition-exemption-4 [https://perma.cc/9EE5-8D23] (Dec. 3, 2021).

221 *See* Itiel E. Dror, Bridget M. McCormack & Jules Epstein, *Cognitive Bias and Its Impact on Expert Witnesses and the Court*, 54 JUDGES' J. 8, 9 (2015) ("[E]xperts are most often recruited by one side of the adversarial system, and work within the team and objectives of that side . . . [which] can subconsciously influence [the expert's] perceptions and judgments.").

222 Thompson, *supra* note 149, at 1522.

research independently will no longer be an impediment to their valid exercise of discretion.

2. Analogous Positions Already in Existence

A scientific advisor is not a wholly novel or radical concept, nor is the creation of a centralized agency to ensure consistent administration of justice through those advisors. Judicial counsel positions are prevalent in courts to assist with conducting research for judges, family law courts often appoint psychologists to evaluate a child's mental state,[223] and law firms working in intellectual property generally hire scientific advisors for patent cases.[224] Despite the recent political trend of downsizing federal agencies,[225] the United States still has an abundance of them,[226] each organized around one central purpose. With the growing prevalence of AI in all facets of life, it seems nearly inevitable that a new agency will need to be created to address the accompanying concerns.

Both the United Kingdom[227] and India[228] have scientific advisors built into their intellectual property and patent law

---

223 *The Judge Appointed a Psychologist in My Divorce Case. Now What?*, DADVOCACY (May 31, 2022), https://dadvocacy.com/blog/2022/05/the-judge-appointed-a-psychologist-in-my-divorce-case-now-what/ [https://perma.cc/JF74-KR33].

224 *See* Lisa Larrimore Ouellette, *Transitioning from Science to Patent Law*, WRITTEN DESCRIPTION (Mar. 15, 2015), https://writtendescription.blogspot.com/2015/03/transitioning-from-science-to-patent-law.html [https://perma.cc/PC2P-8YK7].

225 *See* Elena Shao & Ashley Wu, *The Federal Work Force Cuts So Far, Agency by Agency*, N.Y. TIMES, https://www.nytimes.com/interactive/2025/03/28/us/politics/trump-doge-federal-job-cuts.html [https://perma.cc/P23T-TQ4S] (May 12, 2025).

226 *See The Federal Bureaucracy*, ADELPHI UNIV., https://libguides.adelphi.edu/c.php?g=745658&p=9242744 [https://perma.cc/M468-MYQT] (Aug. 23, 2024) (explaining that there are over 2,000 federal agencies in the United States which, together with the Cabinet departments, employ more than 2.7 million people).

227 *See* Angus Milne & James Simpson, *Patents Court Provides Guidance on Technical Experts vs Scientific Advisers*, HLK (Apr. 4, 2024), https://www.hlk-ip.com/news-and-insights/patents-court-provides-guidance-on-technical-experts-vs-scientific-advisers/ [https://perma.cc/E345-GYBU]. In the United Kingdom, most scientific advisors are appointed to appeals courts, and they are meant to "educate the Court in the relevant technology" that is being offered as evidence, not assist in the determination of any substantive issue. *Id.*

> [T]here are cases at the cutting edge of science, where even the most experienced judges have considered that a non-controversial "teach-in" would be desirable and so the Court . . . has appointed a scientific adviser to assist it with getting to grips with the relevant concepts. Such teach-ins have been more commonplace in the Court of Appeal where it is almost inevitable that at least one [judge] . . . will not be steeped in the relevant science.

Brian Cordery, *The Role of Scientific Advisers in the English Patents Court*, KLUWER PAT. BLOG (Mar. 14, 2024), https://patentblog.kluweriplaw.com/2024/03/14/the-role-of-scientific-advisers-in-the-english-patents-court/ [https://perma.cc/7KNE-WSC3].

system as real-world demonstrations of how such a mechanism can effectively function. The simplest analogy in the United States is court-appointed psychologists in family court. Known as "evaluators," these licensed and trained professionals are ordered by a judge to assess the conditions of a home, determine which custody scenario would be in the best interest of the child, and provide a confidential report.[229] The final custody determination is made by the judge, but the evaluator provides vital information based on their years of experience in the field of mental health.[230] Similarly, the scientific advisor sent by the agency would offer expertise in how to evaluate the validity of various types of machine evidence, but the final decision on admissibility is still reliant on the judge's discretion. And while a psychological evaluator can be cost prohibitive since the parties must bear the expense themselves,[231] the process of providing scientific advisors would already be a component of the legal system available for use.

The cost-prohibitive nature of outside advisors has already negatively impacted defense teams in the context of AI. In the rare case a judge orders disclosure of TrueAllele's source code, Cybergenetics does not make it easy for defendants. As explained by one defense attorney, his client would have had to

---

228 *See* Essenese Obhan & Sayali Gulve, *India: Appointing Scientific Advisors in Patent Disputes*, MONDAQ (Aug. 7, 2020), https://www.mondaq.com/india/patent/973846/appointing-scientific-advisors-in-patent-disputes [https://perma.cc/3VKH-CLAM].

> In India, Section 115 of the Patents Act . . . provides that in any suit for infringement or in any proceeding before a court, the court may at any time . . . appoint an independent scientific adviser, to assist the court or to inquire and report upon any such question of fact or of opinion . . . .
>
> . . . .
>
> A scientific advisor plays a crucial role in educating and presenting intricate technological issues [by] help[ing to] translate complex technology and communicate the legal implications of conclusions into terms the judges and the patent attorneys can understand.

Scientific Advisers in Patent Litigation – The Indian Perspective, IPR STUDIOS, http://iprstudio.com/scientific-advisers-in-patent-litigation-the-indian-perspective/ [https://perma.cc/7AQB-PFC3] (last visited Apr. 18, 2025).

229 *California Courts Self-Help Guide: Child Custody Evaluations*, JUD. BRANCH CAL., https://selfhelp.courts.ca.gov/child-custody/evaluations [https://perma.cc/VQG4-LH4U] (last visited Apr. 9, 2025).

230 *See id.*

231 *See id.*; *see also* Joel S. Seidel et al., *What Is a 730 Custody Evaluation?*, JOEL S. SEIDAL & ASSOCS. (Mar. 15, 2017), https://www.seidellaw.com/blog/2017/march/what-is-a-730-custody-evaluation-and-what-should [https://perma.cc/6VQU-V66J] (explaining that a court-ordered evaluation "can cost anywhere from $1,000 to $100,000," depending on the issues to be assessed).

pay $15,000 to access and review the code.[232] On top of that staggering fee, "the defense expert would also have to obtain $1 million in liability insurance, agree to take only handwritten notes and travel to the company's Pittsburgh headquarters for the review."[233] In all, "it would cost at least $50,000 to comply with the nondisclosure agreement, which also might bar [the] expert witness from testifying."[234] By providing a nationwide mechanism, there will be a fairer system in which all parties can have equal access to AI expertise without going bankrupt.

B. How Would This Work?

While a proposed solution may seem viable or advantageous in the abstract, it must also survive the practicalities of operating within the real world. The central agency would likely need to have a broader purpose covering all areas in which AI is a growing issue, including intellectual property cases and law enforcement's use of AI, but one department would be focused on legal cases involving machine evidence. Parties could submit the evidence at issue to the department, and teams of vetted scientists would analyze the data, prepare a report, and send an advisor to counsel the judge as to the reliability of the evidence being presented. While an alternative method would simply be the use of court-appointed scientific advisors not connected to a centralized agency, concerns of inconsistent analysis and the inability to review millions of lines of code for potential errors without assistance would be assuaged by an agency's more collaborative, yet still confidential, environment.

Throughout the federal court system, there are 94 district courts,[235] and in the fiscal year 2023, there were 68,950 criminal defendant filings.[236] This means, on average, a district court handles over 730 criminal cases per year, discounting differences in jurisdiction that may result in a higher or lower caseload for particular courts. But not all of those cases will necessarily involve machine evidence. TrueAllele claims it has been used as

---

232 Jouvenal, *supra* note 3.

233 *Id.*

234 *Id.*

235 *Introduction to the Federal Court System*, U.S. DEP'T OF JUST., https://www.justice.gov/usao/justice-101/federal-courts [https://perma.cc/7W78-6LP6] (last visited Apr. 7, 2025).

236 *Federal Judicial Caseload Statistics 2023*, U.S. CTS., https://www.uscourts.gov/statistics-reports/federal-judicial-caseload-statistics-2023 [https://perma.cc/APV3-BHQM] (last visited Apr. 8, 2025).

an analysis tool in over 1,000 cases to date,[237] and it was first used at trial in 2009.[238] Therefore, over the last fifteen years, TrueAlelle has been used as evidence at an average rate of 67 cases per year, which is only about 0.1% of those 68,950 yearly filings and less than a single case per district court. Assuming only one case per court, TrueAllele evidence would be present in just 0.1% of the 730 yearly cases. There are over 5,500 Assistant U.S. Attorneys (AUSAs) throughout the nation,[239] so with 68,950 filings in 2023, each AUSA handles on average 12 to 13 criminal cases per year, which constitutes about 2% of the 730 cases assigned to each court. If one AUSA is expected to handle 2% of the yearly caseload for a district court, a team of advisors would be more than capable of handling 0.1% of the same caseload, even given the more arduous and labor-intensive process of analyzing AI for potential falsehoods.

Assuming higher caseloads in certain jurisdictions, more cases involving AI after additional programs like COMPAS are taken into account, and higher rates of machine evidence moving into the future, the system should still be more than adequate. Using 2% as a standard caseload, there is a margin large enough to accommodate a 1,900% increase in the agency's caseload. Because the agency is not involved in substantive fact-finding but is designed to review program source codes as they pertain to particular evidence, the scientific advisors could easily move between multiple cases in a way that attorneys cannot. Furthermore, once the agency gains experience with a certain type of machine evidence and how its reliability should be weighed by judges, efficiency will naturally increase.

As a final safeguard to preserve the integrity of the justice system, the scientists hired by the agency would not only need a requisite degree and background in the scientific field at issue, but they must remain an impartial party not hired by either side nor by the company which owns the program being used. This would require a standardized vetting and certification process as well as the imposition of ethical requirements similar to those already imposed on officers of the court.

---

237 *Demonstrating Our Expertise: Proven Technology*, CYBERGENETICS, https://www.cybgen.com [https://perma.cc/L5BS-HBAS] (last visited Apr. 8, 2025).

238 Jouvenal, *supra* note 3.

239 *NAAUSA Mission*, NAAUSA, https://www.naausa.org/about [https://perma.cc/GHB5-MK2W] (last visited Apr. 8, 2025).

VI. CONCLUSION

Though science fiction writers oft depict futuristic computer intelligences capable of answering questions humanity itself is incapable of solving, the current status of AI is somewhat less illustrious and more so fraught with errors. When those errors manifest in the court room, it places the entire system of justice at risk. As Richard Feynman—the famous American theoretical physicist and Nobel Laureate—once wrote, "What I cannot create, I do not understand."[240] AI can only mimic human creation; it cannot understand, and it cannot be trusted to make objective analyses uncontaminated by bias or human programmer error one hundred percent of the time.

The problem is not necessarily that machine evidence might not always be entirely accurate—courts face that same problem with human experts. The problem lies in the fact that jurors, and potentially even judges, more readily trust computer evidence without questioning it. Hidden algorithms, doubtful validation studies, and proven mistakes are chipping away at the foundation of the legal system. Without the institution of an extra measure of protection, cracks will only continue to form as AI gains further prominence as an evidentiary tool, eventually swallowing the court system whole and leaving defendants' fates in the hands of unseen algorithms. Then, computer programmers will really be "rolling their eyes."[241]

---

240 *Richard Feynman's Blackboard at Time of His Death*, CALTECH ARCHIVES, https://digital.archives.caltech.edu/collections/Images/1.10-29/ [https://perma.cc/5TH6-SZVB] (last visited Apr. 9, 2025).

241 *See* Baker, *supra* note 197 ("Behind every successful AI is a programmer rolling their eyes.").

The *Chapman Law Review* is published by its student members at Chapman University Dale E. Fowler School of Law. The *Chapman Law Review* can be reached online at www.chapmanlawreview.com or by e-mail at chapman.law.review@gmail.com. Submission inquiries can be sent to chapmanlawreview.submissions@gmail.com. The office of the *Chapman Law Review* is located in Donald P. Kennedy Hall on the campus of Chapman University, One University Drive, Orange, CA 92866.