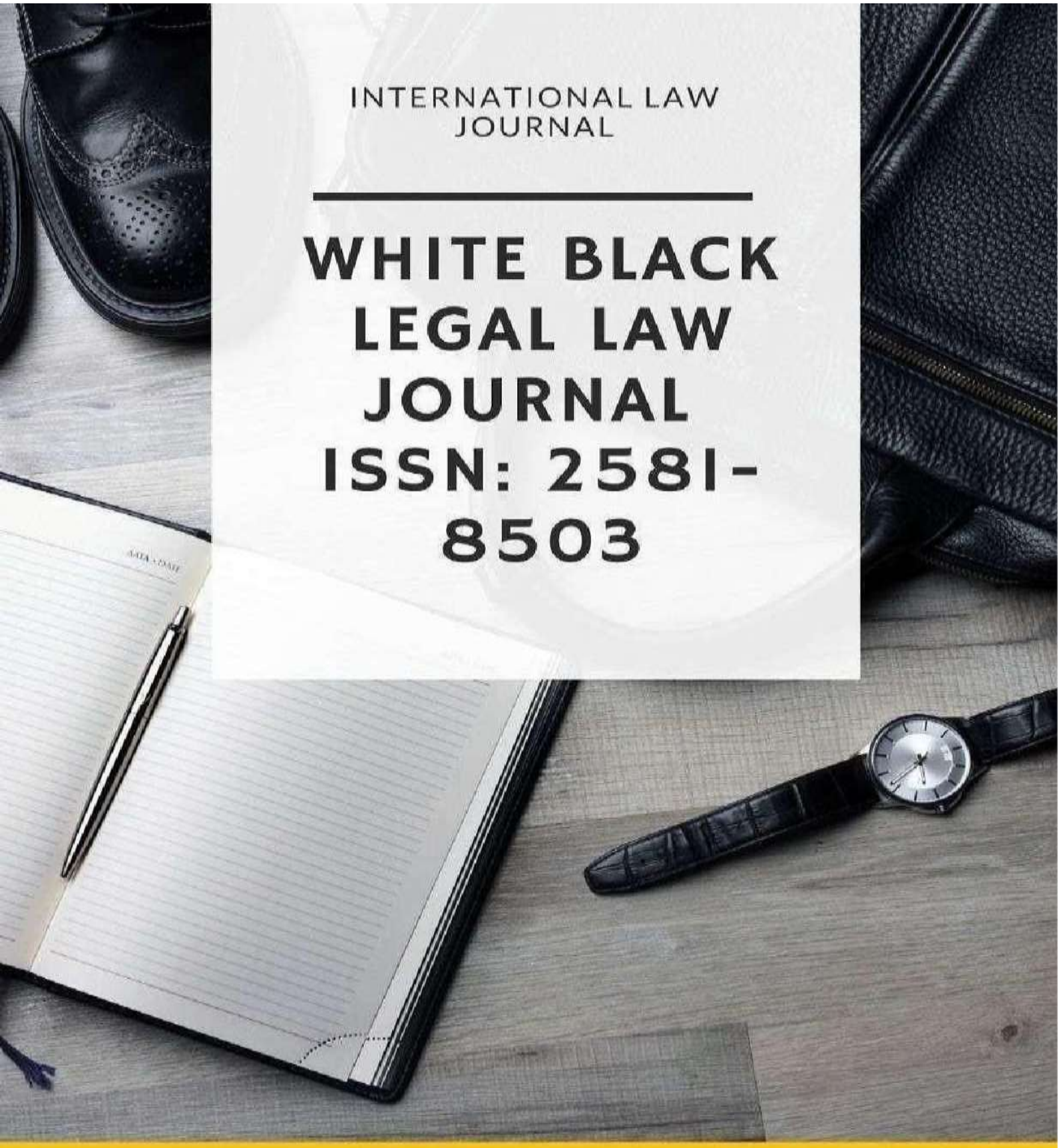

INTERNATIONAL LAW JOURNAL

WHITE BLACK LEGAL LAW JOURNAL

ISSN: 2581-8503

Peer - Reviewed & Refereed Journal

The Law Journal strives to provide a platform for discussion of International as well as National Developments in the Field of Law.

WWW.WHITEBLACKLEGAL.CO.IN

DISCLAIMER

EDITORIAL TEAM

Raju Narayana Swamy (IAS) Indian Administrative Service officer

Dr. Raju Narayana Swamy popularly known as Kerala's Anti-Corruption Crusader is the All India Topper of the 1991 batch of the IAS and is currently posted as Principal Secretary to the Government of Kerala. He has earned many accolades as he hit against the political-bureaucrat corruption nexus in India. Dr Swamy holds a B.Tech in Computer Science and Engineering from the IIT Madras and a Ph. D. in Cyber Law from Gujarat National Law University. He also has an LLM (Pro) (with specialization in IPR) as well as three PG Diplomas from the National Law University, Delhi- one in Urban Environmental Management and Law, another in Environmental Law and Policy and a third one in Tourism and Environmental Law. He also holds a post-graduate diploma in IPR from the National Law School, Bengaluru and a professional diploma in Public Procurement from the World Bank.

Dr. R. K. Upadhyay

Dr. R. K. Upadhyay is Registrar, University of Kota (Raj.), Dr Upadhyay obtained LLB, LLM degrees from Banaras Hindu University & PHD from university of Kota. He has succesfully completed UGC sponsored M.R.P for the work in the Ares of the various prisoners reforms in the state of the Rajasthan.

Senior Editor

Dr. Neha Mishra

Dr. Neha Mishra is Associate Professor & Associate Dean (Scholarships) in Jindal Global Law School, OP Jindal Global University. She was awarded both her PhD degree and Associate Professor & Associate Dean M.A.; LL.B. (University of Delhi); LL.M.; PH.D. (NLSIU, Bangalore) LLM from National Law School of India University, Bengaluru; she did her LL.B. from Faculty of Law, Delhi University as well as M.A. and B.A. from Hindu College and DCAC from DU respectively. Neha has been a Visiting Fellow, School of Social Work, Michigan State University, 2016 and invited speaker Panelist at Global Conference, Whitney R. Harris World Law Institute, Washington University in St. Louis, 2015.

Ms. Sumiti Ahuja

Ms. Sumiti Ahuja, Assistant Professor, Faculty of Law, University of Delhi,
Ms. Sumiti Ahuja completed her LL.M. from the Indian Law Institute with specialization in Criminal Law and Corporate Law, and has over nine years of teaching experience. She has done her LL.B. from the Faculty of Law, University of Delhi. She is currently pursuing PH.D. in the area of Forensics and Law. Prior to joining the teaching profession, she has worked as Research Assistant for projects funded by different agencies of Govt. of India. She has developed various audio-video teaching modules under UGC e-PG Pathshala programme in the area of Criminology, under the aegis of an MHRD Project. Her areas of interest are Criminal Law, Law of Evidence, Interpretation of Statutes, and Clinical Legal Education.

Dr. Navtika Singh Nautiyal

Dr. Navtika Singh Nautiyal presently working as an Assistant Professor in School of law, Forensic Justice and Policy studies at National Forensic Sciences University, Gandhinagar, Gujarat. She has 9 years of Teaching and Research Experience. She has completed her Philosophy of Doctorate in 'Intercountry adoption laws from Uttranchal University, Dehradun' and LLM from Indian Law Institute, New Delhi.

Dr. Rinu Saraswat

Associate Professor at School of Law, Apex University, Jaipur, M.A, LL.M, PH.D,

Dr. Rinu have 5 yrs of teaching experience in renowned institutions like Jagannath University and Apex University. Participated in more than 20 national and international seminars and conferences and 5 workshops and training programmes.

Dr. Nitesh Saraswat

E.MBA, LL.M, PH.D, PGDSAPM
Currently working as Assistant Professor at Law Centre II, Faculty of Law, University of Delhi. Dr. Nitesh have 14 years of Teaching, Administrative and research experience in Renowned Institutions like Amity University, Tata Institute of Social Sciences, Jai Narain Vyas University Jodhpur, Jagannath University and Nirma University.
More than 25 Publications in renowned National and International Journals and has authored a Text book on CR.P.C and Juvenile Delinquency law.

Subhrajit Chanda

BBA. LL.B. (Hons.) (Amity University, Rajasthan); LL. M. (UPES, Dehradun) (Nottingham Trent University, UK); PH.D. Candidate (G.D. Goenka University)

Subhrajit did his LL.M. in Sports Law, from Nottingham Trent University of United Kingdoms, with international scholarship provided by university; he has also completed another LL.M. in Energy Law from University of Petroleum and Energy Studies, India. He did his B.B.A.LL.B. (Hons.) focussing on International Trade Law.

ABOUT US

WHITE BLACK LEGAL is an open access, peer-reviewed and refereed journal provide dedicated to express views on topical legal issues, thereby generating a cross current of ideas on emerging matters. This platform shall also ignite the initiative and desire of young law students to contribute in the field of law. The erudite response of legal luminaries shall be solicited to enable readers to explore challenges that lie before law makers, lawyers and the society at large, in the event of the ever changing social, economic and technological scenario.

With this thought, we hereby present to you

ALGORITHMIC TRADEMARK ENFORCEMENT: EFFICIENCY, ERROR, AND THE CHILLING EFFECT ON LEGITIMATE COMMERCE

AUTHORED BY - VIMALA MARY.A & ASHELLE DEYONA D'SOUZA

ABSTRACT

The digital marketplace is increasingly policed by automated algorithms deployed by online platforms and major rights holders to detect and enforce against potential trademark infringement. While lauded for its speed and scale, this shift towards algorithmic enforcement raises profound legal and ethical concerns. This research critically examines the triple-edged impact of these technologies. It argues that the pursuit of efficiency through automation creates a significant risk of Type I errors false positives where legitimate, non-infringing content is mistakenly removed. Such errors, often arising from the inability of algorithms to contextualize fair use, parody, or comparative advertising, impose a substantial burden on lawful businesses and creators. Furthermore, the opacity and speed of these systems generate a predictable "chilling effect," where actors self-censor expression and commercial activity to avoid the high costs and procedural hurdles of challenging erroneous enforcement actions. This study employs a mixed-methods approach, combining a doctrinal analysis of key legal frameworks like the DMCA and EU's Digital Services Act, a systematic review of platform policies and appeal processes, and a qualitative case study analysis of publicized disputes. The findings will demonstrate that unchecked algorithmic enforcement undermines the balance intrinsic to trademark law, prioritizing the interests of powerful rights holders over competitive markets and free expression. The research concludes by proposing a governance framework for more transparent, accountable, and context-aware enforcement systems that safeguard legitimate commerce.

Keywords: Algorithmic Enforcement, Trademark Law, Chilling Effect, False Positives, Digital Commerce

1. INTRODUCTION

In the past decade, the growth of digital commerce and user-generated content has driven major online platforms to increasingly rely on **automated algorithms** for monitoring, detecting, and removing alleged trademark infringements. These algorithmic enforcement systems promise significant gains in speed and scale: able to scan millions of listings or posts in real time, flag suspicious uses of trademarks, and act swiftly. But this shift comes at a cost. Absent sufficient contextual understanding, automated tools frequently generate **false positives** that is, they remove or block content that in fact is lawful. The consequences for small businesses, creators, and competitors can be severe, including loss of sales, damage to reputation, and curtailed freedom to engage in comparative advertising, parody, or other protected commercial expression.

False positive takedowns are not just isolated errors; they have systemic effects. The procedural burden of challenging an erroneous removal through appeals or litigation tends to favor large rights holders with legal teams, leaving smaller actors with limited resources exposed. Furthermore, the opacity inherent in many enforcement algorithms and notice-and-takedown regimes exacerbates the chilling effect: legitimate actors, fearful of crossing opaque lines or provoking enforcement, may self-censor, avoiding even lawful uses of trademarks. Such chilling undermines the balance that trademark law is meant to strike between protecting brand owners and preserving competition, innovation, and free commercial expression.

While much of the scholarly literature around algorithmic takedowns has focused on **copyright**, less attention has been paid specifically to algorithmic enforcement in trademark contexts. Existing studies of algorithmic detection of trademark infringement are often technical rather than legal-doctrinal, and do not engage as deeply with procedural safeguards, chilling effects, or comparative jurisdictional frameworks. A study examining algorithmic monitoring of intellectual property rights on marketplaces found that algorithms may significantly improve detection speed yet also produce false positives that wrongly flag legitimate listings, highlighting tradeoffs in fairness and legal risk.[1]

[1] A. V. Pokrovskaya, *Intellectual Property Rights Infringement on E-Commerce Marketplaces: Application of AI Technologies, New Challenges*, E3S Web of Conf. 522, 01057 (2024), https://doi.org/10.1051/e3sconf/202452201057.

Legal frameworks like the U.S. *Digital Millennium Copyright Act* (DMCA), though tailored for copyright, include processes (notice-and-takedown, counter-notification) that serve as analogues for trademark enforcement regimes. In the European Union, the *Digital Services Act* (DSA) imposes obligations of transparency and notice for platforms hosting user content, and requires them to put in place mechanisms to contest removal decisions. These frameworks offer both opportunities and challenges: while they provide some procedural safeguards, they do not always ensure that automated systems correctly account for fair use, parody, comparative advertising, or marketplace competition concerns.

This research aims to fill these gaps by doctrinally analyzing how automated trademark enforcement systems operate under the DMCA and the DSA, documenting instances of erroneous takedowns and their impacts, and evaluating whether existing procedural and transparency safeguards are adequate. Ultimately, the goal is to propose a governance model that preserves the efficiencies of algorithmic enforcement while reducing error rates, increasing accountability, and protecting legitimate commerce and expression.

1.1 Research Questions

1. How do the technical limitations of algorithmic enforcement tools lead to erroneous takedowns of legitimate commercial activity, and what is the measurable impact on affected businesses?
2. How do the procedural structures of current notice-and-takedown regimes, coupled with the opacity of algorithms, create a chilling effect that leads actors to self-censor lawful expression and commerce?
3. What legal and technical reforms are necessary to create a more balanced, transparent, and accountable algorithmic enforcement ecosystem that effectively protects trademark rights without stifling legitimate competition and innovation?

1.2 Hypothesis

Algorithmic trademark enforcement, despite its efficiency, produces a high rate of erroneous takedowns that creates a chilling effect, ultimately stifling legitimate commerce and undermining the fundamental balance of trademark law.

2. Literature Review and Theoretical Framework

2.1. The Traditional Balance of Trademark Law: Protection vs. Competition

Trademark law has long been conceptualized not as an absolute property right but as a carefully calibrated mechanism designed to serve distinct public goals. The foundational purpose of trademark protection is to reduce consumer search costs and prevent marketplace confusion about the source or sponsorship of goods and services.[2] As the U.S. Supreme Court articulated in *Qualitex Co. v. Jacobson Products Co.*, trademarks help "assure a producer that it (and not an imitating competitor) will reap the financial, reputation-related rewards associated with a desirable product."[3] This incentive structure is crucial for fostering investment in quality and brand identity. However, this protective function is inherently bounded by an equally critical commitment to preserving robust market competition. Scholars like Mark P. McKenna argue that trademark law's traditional limits, such as the requirement of a likelihood of confusion and the doctrine of functionality, are essential "competition policy ipso facto," preventing trademark rights from being wielded as anti-competitive monopolies over useful product features or broad commercial language.[4]

The equilibrium between protection and competition is therefore the central, fragile axiom of trademark doctrine. This balance is frequently tested when enforcement actions, aimed at protecting a mark, inadvertently suppress legitimate competitive activity. The law tolerates certain uses of another's mark, such as for comparative advertising or descriptive purposes, precisely because this fosters the free flow of commercial information and enables consumers to make informed choices between alternatives.[5] As Barton Beebe elucidates, the modern expansion of trademark rights has often strained this balance, privileging brand owners' control over communicative uses that pose no genuine risk of consumer confusion.[6] The historical tension between exclusive rights and competitive access forms the key backdrop for examining algorithmic enforcement, which risks unsettling this balance by imposing large-scale, context-insensitive protection that tends to ignore pro-competitive safeguards.

[2] William M. Landes & Richard A. Posner, *The Economic Structure of Intellectual Property Law* 166–68 (2003).
[3] *Qualitex Co. v. Jacobson Prods. Co.*, 514 U.S. 159, 163–64 (1995).
[4] Mark P. McKenna, *The Normative Foundations of Trademark Law*, 82 Notre Dame L. Rev. 1839 (2007).. Available at: https://scholarship.law.nd.edu/law_faculty_scholarship/226
[5] J. Thomas McCarthy, *McCarthy on Trademarks and Unfair Competition* § 23:1 (5th ed. 2023).
[6] Barton Beebe, The Semiotic Account of Trademark Doctrine and Trademark Culture, 51 UCLA L. Rev. 621, 628–29 (2004).

2.2. The Doctrine of Fair Use, Parody, and Comparative Advertising in Trademark

To maintain its critical balance, trademark law incorporates specific doctrines that immunize certain unauthorized uses of a mark from liability, even where some consumer confusion might be possible. The statutory "fair use" defense permits the use of another's trademark to describe the user's own goods or services, or to indicate their geographic origin.[7] As courts have emphasized, this classic fair use defense is premised on the understanding that some terms are necessary for effective competition and communication in the marketplace and cannot be wholly appropriated by a single entity.[8] Beyond descriptive fair use, the nominative fair use doctrine, pioneered in cases like *New Kids on the Block v. News America Publishing, Inc.*, allows a defendant to use the plaintiff's mark to refer to the plaintiff's own product for purposes of comparison, criticism, or identification.[9] This doctrine is essential for enabling comparative advertising, product reviews, and compatibility information, all of which are vital to informed consumer choice and market transparency.

The protection of parody highlights how the law accommodates non-commercial expression and social critique, sitting at the crossroads of trademark rights and the First Amendment. For a parody to succeed, it must evoke the original mark to deliver its humorous or critical message, while at the same time making clear that it is not the source of the original product. These context-driven defenses are not accidental loopholes but core safeguards that limit trademark rights, ensuring the law does not suppress competition or free expression. Applying such doctrines demands a nuanced, human understanding of intent, context, and audience, something that automated enforcement systems struggle to replicate.

2.3. Algorithmic Governance: From Copyright to Trademark Enforcement

The scholarly discourse on automated intellectual property enforcement is heavily dominated by the context of copyright, particularly under the notice-and-takedown regime established by the Digital Millennium Copyright Act (DMCA).[10] Seminal work by scholars like Jacqueline D. Lipton and Annemarie Bridy has meticulously documented how automated filters and bots, designed to identify copyright-infringing content, routinely fail to account for the nuances of fair use, leading to the over-removal of lawful expression.[11] This body of literature establishes

[7] 15 U.S.C. § 1115(b)(4) (2018).
[8] KP Permanent Make-Up, Inc. v. Lasting Impression I, Inc., 543 U.S. 111, 122 (2004).
[9] *New Kids on the Block v. News Am. Publ'g, Inc.*, 971 F.2d 302, 308 (9th Cir. 1992).
[10] Digital Millennium Copyright Act, 17 U.S.C. § 512 (2018).
[11] Annemarie Bridy, Copyright Policymaking as Procedural Democratic Process: A Discourse-Theoretic Perspective on ACTA, SOPA, and PIPA, 30 Cardozo Arts & Ent. L.J. 153, 160–61 (2012).

a critical foundation, highlighting the inherent tension between algorithmic efficiency and legal nuance. It reveals a systemic bias towards removal, where the costs of leaving potentially infringing content online are perceived by platforms as far greater than the costs of erroneously removing non-infringing material. This "efficiency bias" creates a procedural framework that is fundamentally skewed against the user.[12]

While this copyright-focused research provides an essential analog, the direct application of its findings to trademark law remains critically underexplored. Trademark enforcement presents a distinct and in some ways more complex challenge for automation. Unlike copyright infringement, which often involves straightforward copying, trademark infringement turns on the inherently contextual "likelihood of confusion" standard, a multifactor test that requires a holistic analysis of consumer perception, market conditions, and intent.[13]

2.4. The "Chilling Effect" in Digital Speech and Commerce: A Theoretical Overview

The chilling effect is a well-established theoretical concept in law, describing the deterrent effect that ambiguous or overly broad laws and enforcement mechanisms have on lawful conduct, where individuals choose to avoid engaging in protected activity for fear of legal repercussions. In the context of free speech, this concept is paramount, as the First Amendment prohibits laws that cast a "pall of orthodoxy" over society by causing speakers to steer far wider of the unlawful zone than the law actually requires.[14] The doctrine has been powerfully extended into the digital realm, where the architecture of enforcement particularly its opacity, automation, and asymmetry can induce significant self-censorship. Scholars like Hannah Bloch-Wehba have argued that automated systems create a "new chilling effect," distinct from its traditional form, because the risk of being targeted is determined by inscrutable technical processes rather than discernible legal standards.[15] This lack of transparency prevents users from understanding what conduct might trigger enforcement, leading them to avoid even plainly lawful behavior.

[12] **Jennifer M. Urban et al.,** *Notice and Takedown in Everyday Practice* **32 (2017) (UC Berkeley Pub. Law Research Paper No. 2755628), Available at:** https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2755628
[13] Restatement (Third) of Unfair Competition § 21 (Am. L. Inst. 1995).
[14] Keyishian v. Bd. of Regents, 385 U.S. 589, 604 (1967)
[15] Hannah Bloch-Wehba, *Global Platform Governance: Private Power in the Shadow of the State*, 72 SMU L. Rev. 27 (2019).

This theoretical framework is directly applicable to commercial expression and competition. The high transactional costs of challenging an erroneous takedown including legal fees, lost revenue during the appeals process, and reputational damage create a powerful economic disincentive for small businesses and creators. As Mark A. Lemley and Eugene Volokh note in the context of copyright, the notice-and-takedown system can too easily be used as a tool of censorship.[16]

2.5. Gaps in the Existing Literature: The Need for a Trademark-Specific Focus

The burgeoning scholarship on algorithmic enforcement has thus far centered predominantly on copyright law and freedom of speech, creating a significant gap in our understanding of its impact on the distinct doctrinal and economic landscape of trademarks. While the theoretical work on the chilling effect and the technical analyses of automated systems provide a crucial foundation, they often treat algorithmic governance as a monolithic force. This approach overlooks the fact that trademark law's core function is not merely to protect a private right but to regulate competition and ensure clarity in the marketplace.[17]

This research seeks to fill this identified void. Existing studies of algorithmic IP enforcement on e-commerce platforms often prioritize engineering perspectives, analyzing detection accuracy rates while engaging only superficially with the legal standards those algorithms are meant to apply.[18]

3. Legal Frameworks Governing Algorithmic Enforcement

3.1. The U.S. Model: Notice-and-Takedown Under the DMCA (as an Analogue)

While the Digital Millennium Copyright Act (DMCA) is a copyright statute, its § 512(c) "notice-and-takedown" regime has become the de facto model for addressing many forms of allegedly infringing user-generated content online, including trademark disputes.[19] This framework provides a safe harbor from monetary liability for platforms that promptly remove content upon receipt of a compliant notice from a rights holder.[20] The system's efficiency for

[16] Mark A. Lemley & Eugene Volokh, *Freedom of Speech and Injunctions in Intellectual Property Cases*, 48 Duke L.J. 147, 150–51 (1998).
[17] McKenna Supra note 4 at 1839, 1841
[18] A. V. Pokrovskaya, *Intellectual Property Rights Infringement on E-Commerce Marketplaces: Application of AI Technologies, New Challenges*, E3S Web of Conf. 522, 01057 (2024), https://doi.org/10.1051/e3sconf/202452201057.
[19] 17 U.S.C. § 512(c) (2018).
[20] *Id.*

rights holders is undeniable; it allows for the rapid removal of content without initial judicial oversight. However, its application to trademark law is profoundly problematic. The DMCA's procedures were designed for copyright infringement, which often involves blatant copying, not for the nuanced, context-driven "likelihood of confusion" analysis required in trademark.[21] This misfit creates a significant risk of error, as a takedown notice need only allege infringement in good faith, placing the burden on the user to file a formal counter-notice to restore their lawful content.[22] This asymmetry inherently pressures platforms to err on the side of removal, directly enabling the over-removal of non-infringing trademark uses.

3.2. The E.U. Model: Transparency and Due Process Under the Digital Services Act (DSA)

The European Union's Digital Services Act (DSA) represents the most significant modern attempt to regulate algorithmic enforcement and create baseline protections for users.[23] Unlike the DMCA, the DSA is horizontally applicable, meaning its rules govern the removal of all types of illegal content, including trademark infringement.[24] The DSA imposes robust transparency obligations on platforms, requiring them to publish detailed reports on their content moderation activities, including the use of automated tools.[25] Crucially, it also mandates the implementation of internal complaint-handling systems and out-of-court dispute resolution mechanisms, providing users with more accessible avenues to challenge takedown decisions than the DMCA's formal counter-notice process.[26] While the DSA does not prohibit algorithmic enforcement, it seeks to mitigate its risks by requiring that very large online platforms conduct annual risk assessments of their systems and allow vetted researchers access to data to study systemic risks.[27] This regulatory approach explicitly acknowledges the chilling effect and other systemic harms of automated moderation, aiming to inject due process and accountability into a previously opaque system.

[21] *Jacqueline D. Lipton, Addressing the Inevitable Disappearance of the DMCA's Notice-and-Takedown System, 19 N.C. J.L. & Tech. 87, 97 (2017).*
[22] 17 U.S.C. § 512(g) (2018).
[23] Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and Amending Directive 2000/31/EC (Digital Services Act), 2022 O.J. (L 277) 1.
[24] *Id.* at art. 1.
[25] *Id.* at art. 15, 42.
[26] *Id.* at art. 17, 20.
[27] *Id.* at art. 26, 34, 40.

3.3. Platform Liability: The Intermediary's Role and Incentives

The legal frameworks of the DMCA and DSA ultimately function by governing the liability and obligations of online intermediaries the platforms that host third-party content.[28] Their design is predicated on a fundamental economic reality: platforms face asymmetric incentives that heavily favor the over-removal of content. Under the DMCA's safe harbor model, a platform's rational choice is to comply with a takedown notice to avoid potentially massive secondary liability litigation, even if the claim appears meritless.[29] The cost of a single lawsuit vastly outweighs the cost of removing a single user's content. This creates a structural bias where the user's interest in maintaining lawful content is systematically undervalued. Furthermore, platforms have an independent interest in maintaining goodwill with major corporate rights holders, who are significant sources of advertising revenue or formal partnership agreements. This commercial pressure can incentivize platforms to develop and deploy aggressive automated enforcement tools, often going beyond legal requirements in what is termed "shadow regulation," to preemptively placate these powerful entities.[30]

The DSA seeks to recalibrate these incentives by imposing affirmative due diligence obligations and transparency requirements, effectively making a platform's efforts to combat illegal content a condition for its liability shield.[31]

4. Mechanisms and Technical Limitations

4.1. Anatomy of an Algorithmic Enforcement System (Technical Mechanisms)

Algorithmic trademark enforcement systems are usually based on machine learning and pattern-recognition technologies, enabling them to function at a scale far beyond human capacity. Their technical design typically follows a multi-stage processing pipeline. The first stage involves **data ingestion and preprocessing**, where the system scans millions of product listings, social media posts, or domain names, converting unstructured text and image data into

[28] Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services (Digital Services Act), 2022 O.J. (L 303) 1.
[29] Digital Millennium Copyright Act of 1998, Pub. L. No. 105-304, 112 Stat. 2860 (codified as amended at 17 U.S.C. §§ 512, 1201–1205 (2018)).
[30] Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 Harv. L. Rev. 1598, 1613 (2018).
[31] Regulation 2022/2065, on a Single Market For Digital Services and Amending Directive 2000/31/EC (Digital Services Act), 2022 O.J. (L 277) 1, art. 1 (EU).

a machine-readable format.[32] The core of the system is the **detection and classification model**. For text, this often relies on natural language processing (NLP) techniques, including keyword matching, n-gram analysis, and semantic similarity models trained to flag listings containing a brand's name or known variations (e.g., "Nike," "Nikee," "N1ke").[33] For image-based detection, convolutional neural networks (CNNs) are trained on a corpus of a brand's official logos and product imagery to identify visual similarities in user-uploaded photos.[34]

4.2. The Context Blind Spot: Why Algorithms Fail at Fair Use and Parody

The fundamental flaw in automated trademark enforcement is the inability of algorithms to comprehend the context that determines whether a use is infringing or lawful. Trademark law's core inquiry is the "likelihood of confusion" that is a multifactor, holistic test that requires an understanding of consumer perception, market structure, and the intent of the user.[35] Defenses like fair use and parody are even more context-dependent, turning on communicative intent, humor, and the transformative nature of the work.[36] Algorithmic systems, by contrast, excel at identifying superficial similarities, a matching word or a visually similar logo but are "legally blind" to the surrounding circumstances that give those signs a different, non-infringing meaning.[37]

This technical limitation directly causes the false positives documented in this research. The algorithm's operation is reductive: it translates the nuanced, open-ended standards of law into a closed set of computationally tractable rules based on similarity scores.[38] This process inevitably results in lawful activity being misclassified as infringing, because the characteristics that make it lawful such as satire, comparison, or descriptive use, fall outside the system's training data and operational scope. The algorithm is designed to detect similarity to a protected mark, not to interpret the legal significance of context, purpose, or consumer perception.

[32] A. V. Pokrovskaya, *Intellectual Property Rights Infringement on E-Commerce Marketplaces: Application of AI Technologies, New Challenges*, E3S Web of Conf. 522, 01057 (2024), https://doi.org/10.1051/e3sconf/202452201057.
[33] *See generally* Christopher D. Manning et al., *Introduction to Information Retrieval* (2008).
[34] Yann LeCun et al., *Deep Learning*, 521 Nature 436, 443 (2015)
[35] Restatement (Third) of Unfair Competition § 21 (Am. L. Inst. 1995).
[36] *See supra* § 2.2.
[37] Hannah Bloch-Wehba, *Automation in Moderation*, 73 Emory L.J. 387, 394 (2023)
[38] Jonathan Zittrain, *The Future of the Internet—And How to Stop It* 125 (2009)

4.3. Case Study Analysis: Documented Instances of False Positives and Their Impact

Theoretical concerns about algorithmic errors translate into tangible, harmful consequences for businesses and creators. Case studies offer clear evidence of the real-world effects of these false positives. A prominent example occurred in 2018, when the automated system on Amazon's Marketplace erroneously removed a multitude of legitimate listings for phone accessories that were described as "compatible with" or "for" a brand-name device like Samsung or Apple.[39] The algorithm, trained to spot the trademarked term, failed to recognize the legal significance of descriptive fair use, treating the mere presence of the mark as evidence of infringement. The impact on the affected third-party sellers was severe, including sudden income loss, lowered seller rankings, and lengthy appeals to reinstate their businesses.

Beyond e-commerce, similar errors occur in content moderation on social media and advertising platforms. Artists and satirists frequently report their accounts being suspended or their content removed when automated systems flag parodic uses of trademarks.[40] For small businesses and individual creators, the time and effort required to contest wrongful removals and the potential for lasting reputational damage can be crippling. These cases shift the conversation from abstract legal theory to concrete economic and expressive harm, showing that the "error rate" of algorithmic systems is more than a statistic; it reflects real-world market disruption and constrained free expression.

4.4. Quantifying the Error: Challenges in Measuring False Positive Rates

Despite the well-documented existence of false positives, a precise, large-scale quantification of their rate in algorithmic trademark enforcement remains elusive. This data gap is not accidental but stems from the profound opacity of the systems in question. Platforms and major rights holders treat the error rates of their proprietary enforcement algorithms as confidential business information, shielding them from public and academic scrutiny.[41] Furthermore, the very architecture of these systems creates a measurement problem: content that is removed is often simply invisible, making it difficult for researchers to compile a representative dataset of takedowns to analyze for error.[42] While some self-reported data exists such as the transparency

[39] Jason Del Rey, *Amazon's Counterfeit Problem is Getting Worse and It's a Prime Headache*, Vox (Feb. 28, 2019, 8:50 AM), https://www.vox.com/2019/2/28/18168354/amazon-marketplace-prime-counterfeit-goods-fake-reviews.
[40] *See, e.g.*, Sarah Burstein, *The Fake™ Trademark Parody Problem*, 92 Wash. L. Rev. 701, 705 (2017).
[41] Hannah Bloch-Wehba, *Global Platform Governance: Private Power, Public Values*, 76 N.Y.U. Ann. Surv. Am. L. 1, 15 (2021)
[42] Nicolas P. Suzor, *Lawless: The Secret Rules That Govern Our Digital Lives* 87 (2019)

reports now mandated under the EU's Digital Services Act, these often report only on the volume of actions taken and appeals received, not on the ground-truth accuracy of the initial decisions.[43]

Consequently, evidence of the problem's scale is often indirect and inferential. Studies of analogous systems, like YouTube's Content ID for copyright, have found significant error rates, suggesting a similar potential for error in trademark contexts.[44] Qualitative evidence, like the case studies in section 4.3 and the growing volume of public complaints from affected sellers and creators, provides a compelling narrative of a systemic issue.[45] The very fact that a definitive error rate cannot be determined is itself a significant finding. It underscores a fundamental accountability gap: without independent auditing or verification of how enforcement systems operate and the outcomes they produce, there is no way to assess whether they comply with legal standards or to hold their operators responsible for the harm caused.

5. The Chilling Effect: From Error to Self-Censorship

5.1. The Procedural Burden: The Imbalance in Challenging Takedowns

The harm of an erroneous algorithmic takedown is compounded by the onerous and asymmetric process required to challenge it. Current notice-and-takedown regimes, modeled on the DMCA, place the entire burden of initiating a redress procedure on the affected user.[46] For a small business or individual creator, the process of filing a formal counter-notice or navigating a platform's opaque appeals portal entails significant transaction costs: lost revenue during the downtime, time spent understanding complex legal procedures, and the potential need to seek legal counsel.[47] This burden creates a strong economic disincentive for challenging even clear-cut errors. By contrast, the rights holder's initial enforcement action is almost cost-free, often completed with a single click through an automated reporting dashboard. This structural imbalance shifts the costs of mistakes onto the least-resourced participants. For small actors facing a takedown, the rational choice is frequently to accept the loss and adjust future behavior

[43] Martin Husovec & Irene Roche Laguna, *Digital Services Act: A Short Primer* 7 (July 5, 2022) (unpublished manuscript), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4154884
[44] Jennifer M. Urban et al., Notice and Takedown in Everyday Practice 32 (UC Berkeley Pub. Law Research Paper No. 2755628, 2017), https://ssrn.com/abstract=2755628.
[45] Amazon.com, *The FTC's Case Against Amazon Would Lead to Higher Prices and Slower Deliveries for Consumers And Hurt Businesses*, About Amazon (Sept. 26, 2023), https://www.aboutamazon.com/news/company-news/amazon-ftc-antitrust-lawsuit-full-response.
[46] 17 U.S.C. § 512(g) (2018).
[47] Jennifer M. Urban et al., *Notice and Takedown in Everyday Practice* 32 (UC Berkeley Pub. Law Research Paper No. 2755628, 2017), https://ssrn.com/abstract=2755628.

to avoid further enforcement, rather than expend limited resources on a lengthy and uncertain appeals process.

5.2. Opacity and Uncertainty: How Lack of Transparency Deters Legitimate Use

The chilling effect is powerfully exacerbated by the opacity that shrouds algorithmic enforcement systems. Users operate under a veil of uncertainty, unable to discern the specific rules governing enforcement or the reasons for a particular takedown decision.[48] This opacity manifests in two key ways: first, in the secret and proprietary nature of the algorithms themselves, whose internal logic and training data are treated as trade secrets; and second, in the generic and uninformative nature of the notifications users receive, which typically cite a broad violation of "intellectual property policies" without specifying the accused content or the nature of the alleged infringement.[49] This lack of explainability prevents users from understanding what behavior triggered the action, making it impossible to adjust their conduct in a precise and lawful manner. Instead, they must "steer far wider of the unlawful zone," avoiding not only potentially infringing uses but also entire categories of lawful expression such as comparative advertising or parody that they perceive as risky.[50]

This preventive self-censorship lies at the heart of the chilling effect. When the rules are opaque and the consequences of error severe, the most rational strategy is often non-participation. A seller may avoid using a competitor's trademark, even for truthful descriptive purposes protected under fair use. An artist might refrain from creating satirical works featuring well-known logos to sidestep the risk of account suspension.

6. Conclusion, Findings, and a Framework for Reform

6.1. Synthesis of Key Findings

This research has demonstrated that the adoption of algorithmic enforcement in trademark law creates a triple-edged threat: it is efficient yet error-prone, opaque, and systemically chilling. The technical analysis revealed that these systems are inherently context-blind, structurally incapable of applying the nuanced, fact-specific standards of trademark law, particularly the defenses of fair use and parody. The legal examination confirmed that existing frameworks like the DMCA and DSA, while differing in approach, are insufficient to mitigate these flaws; the

[48] Hannah Bloch-Wehba, *Automation in Moderation*, 73 Emory L.J. 387, 394 (2023).
[49] Nicolas P. Suzor, *Lawless: The Secret Rules That Govern Our Digital Lives* 87 (2019).
[50] Keyishian v. Bd. of Regents, 385 U.S. 589, 604 (1967)

DMCA's procedural asymmetry and the DSA's nascent transparency requirements do not compel the accuracy or explainability needed to prevent widespread error. The consequence, as documented through case studies and theoretical analysis, is a significant chilling effect where lawful businesses and creators self-censor to avoid the high costs and uncertainties of challenging automated decisions. This undermines the foundational balance of trademark law, stifling the very competition and expression the doctrine is meant to protect.

6.2. Proposed Principles for a Governance Model

To realign algorithmic enforcement with legal principles, a new governance model based on the following principles is necessary:

6.3 Enhanced Transparency and Explainability: Platforms should be required to provide detailed, specific reasons for takedowns, identifying the accused content and the specific trademark right claimed. Very large platforms should be subject to independent audits of their algorithmic systems to assess error rates and bias.

6.3.1 Robust Human Review and Proportionality: Automated decisions should never result in the most severe sanctions (e.g., account termination) without mandatory human review. Enforcement actions must be proportionate, with temporary holds preferred over immediate, permanent removals during disputes.

6.3.2 Reformed Appeal and Counter-Notice Mechanisms: Appeal processes must be low-cost, rapid, and user-friendly. The burden of proof should shift to the rights holder to substantiate a claim if a user contests a takedown, reversing the current incentive structure.

6.3.3 Algorithmic Training for Context-Awareness: While a technical challenge, investment in AI training sets and models that incorporate examples of lawful fair use, parody, and comparative advertising is essential to reduce the root cause of errors.