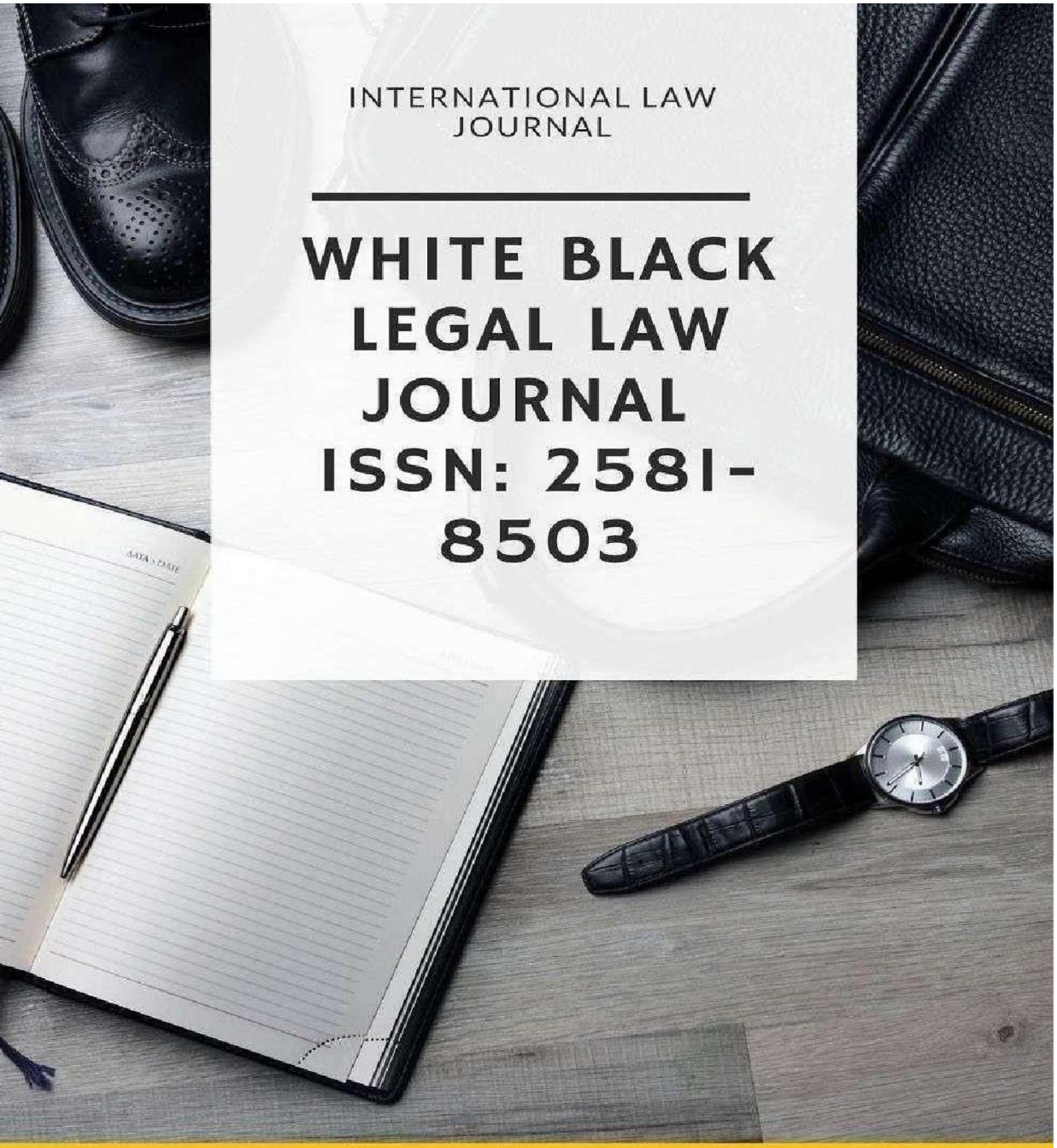

INTERNATIONAL LAW JOURNAL

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

ABOUT WHITE BLACK LEGAL

White Black Legal – The Law Journal is an open-access, peer-reviewed, and refereed legal journal established to provide a scholarly platform for the examination and discussion of contemporary legal issues. The journal is dedicated to encouraging rigorous legal research, critical analysis, and informed academic discourse across diverse fields of law.

The journal invites contributions from law students, researchers, academicians, legal practitioners, and policy scholars. By facilitating engagement between emerging scholars and experienced legal professionals, *White Black Legal* seeks to bridge theoretical legal research with practical, institutional, and societal perspectives.

In a rapidly evolving social, economic, and technological environment, the journal endeavours to examine the changing role of law and its impact on governance, justice systems, and society. *White Black Legal* remains committed to academic integrity, ethical research practices, and the dissemination of accessible legal scholarship to a global readership.

AIM & SCOPE

The aim of *White Black Legal – The Law Journal* is to promote excellence in legal research and to provide a credible academic forum for the analysis, discussion, and advancement of contemporary legal issues. The journal encourages original, analytical, and well-researched contributions that add substantive value to legal scholarship.

The journal publishes scholarly works examining doctrinal, theoretical, empirical, and interdisciplinary perspectives of law. Submissions are welcomed from academicians, legal professionals, researchers, scholars, and students who demonstrate intellectual rigour, analytical clarity, and relevance to current legal and policy developments.

The scope of the journal includes, but is not limited to:

- Constitutional and Administrative Law
- Criminal Law and Criminal Justice
- Corporate, Commercial, and Business Laws
- Intellectual Property and Technology Law
- International Law and Human Rights
- Environmental and Sustainable Development Law
- Cyber Law, Artificial Intelligence, and Emerging Technologies
- Family Law, Labour Law, and Social Justice Studies

The journal accepts original research articles, case comments, legislative and policy analyses, book reviews, and interdisciplinary studies addressing legal issues at national and international levels. All submissions are subject to a rigorous double-blind peer-review process to ensure academic quality, originality, and relevance.

Through its publications, *White Black Legal – The Law Journal* seeks to foster critical legal thinking and contribute to the development of law as an instrument of justice, governance, and social progress, while expressly disclaiming responsibility for the application or misuse of published content.

DEFAMATION VIA EMAIL AND MESSAGING PLATFORMS: LEGAL IMPLICATIONS IN THE DIGITAL AGE

AUTHORED BY - R. PRIYANGA
LLM Candidate (Cyber Law & Security)
SRM Institute of Science & Technology

ABSTRACT

Defamation in the digital era has evolved into a multi-jurisdictional legal challenge of considerable complexity. The instantaneous and borderless character of electronic communication encompassing email, end-to-end encrypted messaging applications, and group-chat platforms has dramatically amplified the speed and geographic reach with which false, misleading, or reputationally harmful statements may propagate. This Article examines the legal implications of defamation transmitted through email and messaging platforms, synthesizing landmark case law, comparative statutory frameworks, and evolving theories of platform liability across the United States, United Kingdom, European Union, India, Australia, Singapore, and Canada. Building upon foundational scholarship in defamation law, the Article identifies three structural tensions that contemporary doctrine must resolve: (1) the public–private communication binary and its erosion in group-messaging environments; (2) the jurisdictional indeterminacy produced by the single-publication rule's migration to borderless networks; and (3) the liability gap created by intermediary immunity regimes. The Article further assesses emerging challenges posed by artificial-intelligence-generated content, deepfakes, and anonymous or pseudonymous authorship. It concludes with a normative framework recommending harmonised international standards, graduated intermediary liability, and technology-assisted content moderation calibrated to protect free expression while providing effective remedies for defamation victims.

I. INTRODUCTION

Defamation has long occupied a foundational position in the private law of civil wrongs, traditionally anchored in the printing press, the broadcast studio, and the public square. The rapid proliferation of digital communication technologies has, however, fundamentally

disrupted these spatial and temporal coordinates.[1] Emails and instant-messaging applications WhatsApp, Telegram, Signal, Slack, and Microsoft Teams, among others now constitute indispensable conduits for personal, professional, and political discourse. These platforms combine the permanence of written libel with the velocity of viral dissemination, creating a risk profile that existing legal frameworks are only beginning to address.

The doctrinal novelty of digital defamation is not merely technological. The architecture of modern messaging platforms challenges each classical element of a defamation claim: 'publication' occurs across multiple jurisdictions simultaneously identification may be effected through algorithmic aggregation of contextual data; 'harm' is amplified and archived indefinitely and 'fault' must be assessed against the conduct not only of natural persons but also of platform intermediaries and, increasingly, autonomous AI systems.[2]

This Article proceeds in seven Parts. Part II presents a comprehensive literature review situating the Article within existing scholarship. Part III defines defamation and traces the evolution of its core elements in digital environments. Part IV examines the specific characteristics of email- and messaging-platform defamation, including employer liability and encrypted-messaging challenges. Part V surveys legal consequences and remedies across key jurisdictions. Part VI addresses jurisdictional conflicts and international law. Part VII proposes prevention strategies and examines the role of AI-driven content moderation. Part VIII offers conclusions and policy recommendations.

II. LITERATURE REVIEW

Eric Barendt's *Freedom of Speech* provides the foundational doctrinal framework within which reputation and expression are balanced, arguing that defamation law must be construed narrowly to avoid chilling legitimate public discourse.[3] Barendt's comparative analysis of British and American approaches reveals a persistent structural divergence: Anglo-Commonwealth systems place the burden of proof on defendants to establish truth, whereas the United States, following *New York Times Co. v. Sullivan,* imposes an 'actual malice' threshold that privileges robust public debate.[4]

Matthew Collins's *The Law of Defamation and the Internet* remains the leading practitioner text on internet defamation.[5] Collins systematically maps defamation doctrine onto online

publication, addressing the single-publication rule, the identification requirement in pseudonymous contexts, and the liability of internet service providers (ISPs). His treatment of 'publication' as a jurisdictional trigger anticipates the central holding of the Australian High Court in *Dow Jones & Co. v. Gutnick,* which established that publication occurs where defamatory content is downloaded and comprehended, not merely where it is uploaded.[6]

Daniel Solove's *The Future of Reputation* examines defamation and privacy through the lens of network theory, demonstrating that the viral dynamics of digital communication generate reputational harm at a scale and speed uncontemplated by pre-digital tort law.[7] Solove's socio-legal analysis informs this Article's treatment of the 'virality factor' in messaging-platform defamation.

The question of whether, and to what degree, digital platforms bear responsibility for user-generated defamatory content has generated an extensive literature. David Ardia's influential article *Reputation in a Networked World* argues that social-foundation theories of defamation must be revisited to account for intermediary architecture, contending that platforms function as reputation-affecting actors even when they do not originate defamatory content.[8] Ardia's framework underpins this Article's analysis of graduated intermediary liability.

Olivier Sylvain has proposed 'intermediary design duties' as a normative category that would impose affirmative obligations on platforms to architect their services in ways that reduce defamatory harm, rather than merely requiring post hoc content removal.[9] This proposal has particular relevance to encrypted-messaging applications, where the technical impossibility of content scanning creates a regulatory vacuum that national courts have begun to fill through group-administrator liability doctrines.

The European approach to intermediary liability has evolved significantly with the Digital Services Act 2022 (DSA),[10] which imposes notice-and-action obligations on 'very large online platforms' and establishes a risk-based regulatory tier. Lilian Edwards and Charlotte Waelde's *Law and the Internet* provides comprehensive doctrinal analysis of EU intermediary liability, tracing the evolution from the E-Commerce Directive's 'mere conduit' framework to the DSA's proactive risk-management obligations.[11]

The multi-jurisdictional dimensions of online defamation have attracted sustained scholarly

attention. Lyrissa Barnett Lidsky's pioneering article *Silencing John Doe: Defamation and Discourse in Cyberspace* identified, at an early stage of the internet's development, the destabilising effect that online anonymity would have on traditional defamation enforcement mechanisms.[12] Lidsky foresaw the tension between the speaker's interest in anonymous self-expression and the defamed party's interest in identifying and pursuing her tormentor a tension that remains unresolved in most legal systems.

Jeffrey Rosen's analysis of the right to be forgotten examines the European Court of Justice's landmark decision in *Google Spain SL v. Agencia Española de Protección de Datos,* which recognised a data subject's right to request de-listing of search results containing defamatory or outdated personal information.[13] Rosen critiques the decision's ambiguous scope and its potential conflict with freedom of expression, a tension that the GDPR subsequently attempted to resolve through Article 17's qualified right to erasure.[14]

Frank Pasquale's *The Black Box Society* situates online reputation management within a broader critique of algorithmic opacity, arguing that search engines and social-media platforms exercise quasi-governmental power over individuals' digital identities without corresponding accountability.[15] This critique informs this Article's policy recommendations regarding AI-generated defamation and the need for algorithmic transparency obligations.

Despite the breadth of scholarship canvassed above, several lacunae remain. First, existing literature has devoted limited attention to defamation occurring in *private or semi-private* messaging environments group chats, workplace channels, and encrypted communications as distinct from the public-facing social-media posts that have dominated case law. Second, the liability of group-chat administrators in common-law jurisdictions remains undertheorised. Third, the intersection of AI-generated content and defamation doctrine has received only preliminary academic treatment. This Article seeks to address these gaps through comparative doctrinal analysis and normative reconstruction.

III. UNDERSTANDING DEFAMATION IN THE DIGITAL CONTEXT

A. Definition and Essential Elements

Defamation, in its classical formulation, comprises any false statement of fact that is published to a third party, identifies the plaintiff, and causes reputational harm.[16] In most common-law

jurisdictions five elements must be established: (1) a false statement of fact; (2) publication to at least one third party; (3) identification of the plaintiff; (4) reputational harm; and (5) the requisite degree of fault negligence for private figures, and actual malice for public figures under United States doctrine.[17]

Each element acquires distinctive features in the digital environment. The 'publication' element is perhaps most significantly transformed: in digital contexts, a single act of transmission may constitute simultaneous publication across hundreds of jurisdictions. The 'identification' requirement may be satisfied not by naming the plaintiff but by providing contextual details sufficient to enable members of a plaintiff's community to draw the defamatory inference. And the 'harm' requirement traditionally assessed by reference to damage to standing in a community must now account for the indefinite archival of defamatory content and its indexation by search engines.

B. Libel Versus Slander in Digital Communication

The libel–slander dichotomy, which traditionally distinguished permanent written defamation from transient spoken defamation, has been substantially complicated by digital media. Written emails and chat-log messages plainly constitute libel.[18] Voice notes and video messages present a harder categorisation problem: while recorded in a permanent medium, they replicate the conversational cadence of speech. Courts in the United Kingdom, following the reasoning in *Monroe v. Hopkins,* have treated digital messages as libel on the basis of their permanence and shareability, irrespective of their conversational register.[19]

The UK Supreme Court's decision in *Stocker v. Stocker* introduced a 'reasonable reader' standard calibrated to the informal register of online communication, holding that courts must interpret digital statements as an ordinary social-media user would understand them rather than applying a strict literalist construction.[20] This contextual interpretive approach represents a significant doctrinal innovation with implications for the assessment of defamatory meaning in messaging-platform communications.

C. Evolution of Defamation Law in the Digital Era

Defamation law has been incrementally adapted to digital media through a combination of legislative reform and judicial development. The United Kingdom's Defamation Act 2013 introduced a 'serious harm' threshold, requiring plaintiffs to demonstrate that the statement has caused, or is likely to cause, serious harm to their reputation a requirement designed to filter

out trivial claims generated by the volume of online communication.[21]

India has approached digital defamation through a dual civil and criminal regime. Sections 499 and 500 of the Indian Penal Code criminalise defamatory imputations, and the Supreme Court of India upheld these provisions in *Subramanian Swamy v. Union of India,* holding that the protection of reputation is a constitutionally cognisable value that may justify reasonable restrictions on free expression.[22] The Court's endorsement of criminal defamation in the digital context has been widely criticised by freedom-of-expression advocates but reflects a broader global pattern in which reputational protection is treated as a competing fundamental right.[23]

D. Selected Case Studies

Delfi AS v. Estonia (2015) established at the European level that online platforms may be held liable for third-party defamatory comments if they fail to take adequate and timely remedial action, even in the absence of direct authorship.[24] The Grand Chamber of the European Court of Human Rights held that the imposition of liability on Delfi AS for anonymous user comments was compatible with Article 10 of the European Convention on Human Rights, given the commercial nature of the platform, the seriously harmful character of the comments, and the portal's failure to act promptly upon notification.

Lee Hsien Loong v. Leong Sze Hian (2020) established in Singapore that the mere act of sharing defamatory content on social media without endorsement or comment constitutes actionable publication, because sharing materially contributes to the spread of falsehood.[25] The decision has significant implications for group-messaging environments in which forwarding of defamatory material is routine.

IV. DEFAMATION VIA EMAIL AND MESSAGING PLATFORMS

A. The Public–Private Communication Spectrum

The classical defamation requirement of 'publication to a third party' presupposes a binary between private communications (addressed solely to the subject) and public broadcasts. Digital messaging collapses this binary.[26] One-to-one private emails and direct messages may technically satisfy the publication requirement if a defamatory statement is communicated to even a single third party. More significantly, group chats particularly large WhatsApp or Telegram groups whose membership extends beyond a circle of intimates may constitute the functional equivalent of a public forum for publication purposes.

Courts have begun to distinguish between 'genuinely private' communications shared within relationships of trust and 'semi-public' group communications whose reach is indistinguishable from a limited broadcast. The permanent record effect of digital communications the fact that emails and chat histories may be retrieved, screenshotted, and redistributed indefinitely further undermines arguments for treating messaging-platform communications as transient private speech.[27]

B. Legal Framework Governing Email and Instant Messaging

In the United States, Section 230 of the Communications Decency Act provides broad immunity to platform providers for third-party content, shielding email services and messaging applications from defamation liability arising from user communications.[28] This immunity does not, however, protect the individual authors of defamatory messages. In the European Union, the Digital Services Act 2022 imposes graduated obligations on platforms proportionate to their systemic risk, including notice-and-action obligations and, for very large platforms, proactive risk-assessment requirements.[29]

In India, the Information Technology Act 2000 and its 2021 Intermediary Guidelines impose due-diligence obligations on 'significant social media intermediaries,' requiring the appointment of grievance officers, the establishment of takedown mechanisms, and controversially the traceability of encrypted message originators. The traceability obligation has been challenged before Indian courts as incompatible with the right to privacy and end-to-end encryption.[30]

C. Employer–Employee Liability in Workplace Communications

Workplace communication platforms Slack, Microsoft Teams, and corporate email systems create distinctive defamation risks at the intersection of employment law and tort. An employer may incur vicarious liability for defamatory statements made by employees in the course of their employment, including statements transmitted via workplace messaging platforms. The critical question is whether the defamatory communication occurred within the scope of employment a contextual assessment that turns on factors including the purpose of the communication, the platform on which it was made, and the relationship between the statement's subject matter and the employee's professional role.

Courts have held that 'reply-all' emails and mass internal communications that falsely impugn a colleague's professional competence or personal character may constitute actionable libel.

Employers are accordingly advised to implement communication policies that clearly define permissible uses of workplace platforms and establish internal grievance mechanisms as an alternative to public defamatory complaint.

D. Encrypted Messaging, Group Chats, and Anonymous Communications

The end-to-end encryption offered by WhatsApp, Signal, and Telegram's 'Secret Chats' function presents significant challenges for defamation enforcement. Encryption renders content inaccessible to platform providers, limiting their capacity to moderate or remove defamatory material. Courts in India, Singapore, and Australia have, however, held that group-chat administrators may bear independent liability for defamatory content posted by group members if they fail to act upon notification a principle that has been contested on the ground that administrators lack the technical ability to decrypt and pre-screen content.[31]

Anonymous and pseudonymous authorship compounds these difficulties. Courts have adopted subpoena mechanisms to compel platforms to disclose the identity of anonymous defamers. In *Google LLC v. Equustek Solutions Inc.,* the Supreme Court of Canada affirmed the authority of courts to order global de-indexing of defamatory content, establishing a precedent for extraterritorial injunctive relief against platform intermediaries.[32]

V. LEGAL CONSEQUENCES AND REMEDIES

A. Civil Versus Criminal Liability

The majority of common-law jurisdictions classify defamation as a civil wrong, providing monetary compensation and injunctive relief as primary remedies. Civil liability requires proof, on the balance of probabilities, of the essential elements of defamation. In the United States, the fault threshold varies by plaintiff category: public officials and public figures must demonstrate 'actual malice' knowledge of falsity or reckless disregard for the truth while private figures need only establish negligence.[33]

Criminal defamation provisions persist in India, Thailand, and several other jurisdictions, providing an additional layer of deterrence but frequently attracting criticism as instruments of suppression of legitimate political speech. The Supreme Court of India's endorsement of criminal defamation in *Subramanian Swamy v. Union of India* has been noted with concern by the UN Special Rapporteur on Freedom of Expression, who has recommended that criminal defamation laws be repealed or substantially reformed to align with international human rights standards.[34]

B. Remedies: Damages, Injunctions, and Takedown Orders

Civil defamation remedies encompass compensatory damages for actual loss, general damages for reputational harm and emotional distress, and punitive damages where the defendant acted with malice. The quantum of damages in high-profile digital defamation cases has increased substantially: in *Monroe v. Hopkins,* the UK High Court awarded £24,000 in damages for a series of defamatory tweets, while in *Depp v. Heard* (2022), a Virginia jury awarded compensatory and punitive damages of $15 million for false allegations published across multiple media platforms.[35]

Injunctive relief ordering the removal of defamatory content has assumed increasing practical importance given the permanence of digital publication. Courts have exercised jurisdiction to grant mandatory injunctions requiring platform providers to de-list or remove defamatory content, building on the principle established in *Google Spain* and codified in the GDPR's Article 17 right to erasure.[36]

C. Defences to Defamation Claims

Three primary defences are available to defendants in defamation proceedings. First, the defence of *truth* (or 'justification' in pre-2013 UK terminology) is an absolute defence in all common-law jurisdictions: a true statement, however damaging, cannot constitute actionable defamation. The burden of proving truth rests with the defendant.

Second, *privilege* protects certain communications from defamation liability. Absolute privilege attaches to statements made in judicial proceedings, legislative debates, and official governmental communications. Qualified privilege attaches to communications made in good faith to persons with a legitimate interest in receiving them, subject to disproof of malice. The Reynolds privilege for responsible journalism, recognised by the UK House of Lords in *Reynolds v. Times Newspapers Ltd,* extends qualified protection to media publications meeting standards of editorial responsibility, and has been codified in the Defamation Act 2013 as the 'public interest' defence.[37]

Third, the defence of *honest opinion* (formerly 'fair comment') protects expressions of opinion that are based on true facts and are recognisable as opinion rather than fact. In *Gertz v. Robert Welch, Inc.,* the United States Supreme Court held that pure expressions of opinion are categorically protected by the First Amendment and cannot constitute defamation irrespective of their reputational impact.[38]

D. Platform Liability and Content Moderation Obligations

The liability of platform intermediaries for defamatory user content remains one of the most contested areas of digital law. The United States' Section 230 immunity regime has been characterised as the foundation of the open internet, shielding platforms from defamation claims and enabling user-generated content to flourish.[39] Critics, however, argue that the immunity regime has enabled platforms to profit from defamatory content while evading accountability for its harms.

The Australian court's decision in *Google Inc. v. Duffy* represents a significant departure from the US immunity model, holding Google liable as a publisher of defamatory search-result snippets on the grounds that its algorithmic curation constituted a form of publication.[40] This decision illustrates the divergence between the US and Commonwealth approaches to platform liability that continues to frustrate efforts at international harmonisation.

VI. JURISDICTIONAL CHALLENGES AND INTERNATIONAL LAW

A. Cross-Border Defamation and the Single-Publication Rule

The multi-jurisdictional character of online publication creates acute difficulties in identifying the governing law of a defamation claim. The traditional single-publication rule, developed in the context of newspaper circulation, permits only one cause of action arising from a mass publication regardless of the number of copies distributed. Courts in the United States have extended this rule to online publications, limiting defendants' exposure to a single limitation period.

Australian and English courts have declined to adopt the single-publication approach without legislative endorsement. The Australian High Court in *Dow Jones & Co. v. Gutnick* held that the publication of an online article constituted a separate and actionable defamation in each jurisdiction in which the article was downloaded and comprehended a ruling with profound implications for defendants who publish globally.[41] The Defamation Act 2013 subsequently introduced a single-publication rule in England and Wales, starting the limitation period from the date of first publication rather than each subsequent access.[42]

B. Forum Shopping and Libel Tourism

The divergence in national defamation standards has generated the practice of 'libel tourism' the selection by plaintiffs of plaintiff-friendly jurisdictions in which to bring defamation proceedings, irrespective of meaningful connection between the jurisdiction and the

publication or the parties. The United Kingdom, prior to the Defamation Act 2013's serious-harm threshold, was a favoured destination for libel tourism, attracting claims from foreign plaintiffs based on minimal domestic publication. *Rachel Ehrenfeld v. Mahfouz* illustrated the consequences of this phenomenon for US authors, prompting Congress to enact the SPEECH Act 2010, which bars the enforcement of foreign defamation judgments inconsistent with First Amendment standards.[43]

C. Conflict of Laws: Choice of Applicable Law

The choice-of-law question in cross-border defamation turns on competing connecting factors: the place of publication, the place of damage, the domicile of the parties, and the location of the platform's servers. The EU's Rome II Regulation selects the law of the country in which the damage occurs as the default rule, with an exception where the parties share a common habitual residence. Courts have found this rule difficult to apply to online defamation given the global distribution of damage.

The CJEU's decision in *Google v. CNIL* addressed the extraterritorial scope of the right to be forgotten, holding that while EU member-state data protection authorities may order de-listing within the EU, they may not, as a matter of EU law, require global de-listing a ruling that reflects the Court's sensitivity to the territorial limits of EU regulatory competence.[44] Conversely, *Yahoo! Inc. v. LICRA* illustrated the risk of regulatory overreach when a French court ordered Yahoo to remove Nazi-related content globally, generating a US constitutional confrontation that the Ninth Circuit ultimately resolved in Yahoo's favour.[45]

D. International Human Rights Framework

International human rights law recognises both freedom of expression (Article 19 ICCPR) and the right to privacy and reputation (Article 17 ICCPR) as fundamental rights requiring positive protection by states.[46] The Human Rights Committee's General Comment No. 34 acknowledges that defamation laws may constitute permissible restrictions on freedom of expression under Article 19(3), provided that they are provided by law, pursue a legitimate aim, and are demonstrably necessary and proportionate. The Committee has, however, expressed concern that criminal defamation provisions are frequently deployed as instruments of political suppression rather than genuine reputation protection.

VII. PREVENTION, RISK MITIGATION, AND THE ROLE OF ARTIFICIAL INTELLIGENCE

A. Best Practices for Individuals and Organisations

Effective defamation risk management begins with individual responsibility. Individuals should verify the factual basis of statements before transmission, exercise caution in group-messaging environments, and familiarise themselves with the defamation laws of their jurisdiction. The risks associated with forwarding unverified content are particularly acute: as *Lee Hsien Loong v. Leong Sze Hian* demonstrates, passive sharing of defamatory content may attract liability equivalent to original authorship.[47]

Organisations should adopt comprehensive digital communication policies that define permissible and impermissible uses of workplace platforms, establish internal complaint-handling mechanisms, and provide regular training on defamation risk. Non-disparagement clauses in employment contracts and social-media policies can reduce, though cannot eliminate, the risk of employee-generated defamation. Brand-monitoring technologies that track online mentions in near-real time enable early intervention before defamatory content achieves viral traction.

B. Platform Governance and Ethical Content Moderation

Digital platforms occupy a quasi-regulatory position in the defamation ecosystem, exercising private governance over vast quantities of user-generated content. Responsible platform governance requires transparent content-moderation policies, effective notice-and-action mechanisms, and meaningful procedural safeguards for users whose content is subject to removal. The DSA's requirements for transparent moderation, appeals mechanisms, and out-of-court dispute settlement represent a significant advance in this direction.[48]

C. Artificial Intelligence in Content Moderation and AI-Generated Defamation

AI-driven natural language processing tools are increasingly deployed by platforms to identify and flag potentially defamatory content at scale. Sentiment analysis, named-entity recognition, and contextual classification algorithms can detect defamatory patterns with increasing accuracy, enabling automated or assisted moderation that reduces the burden on human reviewers. These tools are, however, subject to errors of both over-inclusion (removing legitimate expression) and under-inclusion (failing to detect sophisticated defamatory content). The emergence of AI-generated defamatory content particularly 'deepfake' audio and video

fabrications presents a qualitatively new legal challenge. Existing defamation doctrine requires identification of a human author AI-generated content may be produced without meaningful human authorship in the traditional sense. This Article submits that legislative frameworks should impose liability on developers and deployers of generative AI systems for foreseeable defamatory outputs, subject to reasonable safeguards, analogous to product liability principles.[49]

VIII. CONCLUSION AND POLICY RECOMMENDATIONS

Defamation via email and messaging platforms constitutes one of the most pressing and under addressed challenges in contemporary private law. The doctrinal frameworks developed for print and broadcast media are structurally inadequate to address the velocity, permanence, anonymity, and multi-jurisdictionality of digital defamation. Courts around the world have made significant progress in adapting defamation doctrine to digital realities, but legislative reform and international cooperation are essential to address the systemic gaps that ad hoc judicial development cannot fill.

This Article advances six policy recommendations. *First,* national legislatures should adopt serious-harm thresholds calibrated to the specific harms of digital publication, drawing on the model of the UK Defamation Act 2013 while expanding its scope to encompass private and semi-private messaging environments. *Second,* intermediary liability regimes should adopt a graduated model that calibrates obligations to platform architecture and risk, moving beyond both the US immunity model and the EU's one-size-fits-all approach. *Third,* group-messaging administrator liability should be codified on a knowledge-and-failure-to-act basis, with clear safe-harbour protections for administrators who act promptly and in good faith. *Fourth,* international cooperation frameworks should establish harmonised conflict-of-laws rules for online defamation, drawing on the model of the SPEECH Act 2010 to resist jurisdictional overreach while providing effective remedies for cross-border victims. *Fifth,* legislative frameworks should address AI-generated defamation through developer and deployer liability provisions modelled on product liability principles. *Sixth,* criminal defamation provisions should be repealed or substantially reformed to conform to international human rights standards, reserving the criminal sanction for the most serious cases involving calculated campaigns of reputational destruction.

The digital ecosystem demands legal frameworks that are simultaneously adaptive to technological change and anchored in enduring principles of reputational dignity, free expression, and procedural fairness. The architecture of this framework must be built now, before AI-generated defamation and ubiquitous encrypted messaging render the existing patchwork of national laws entirely obsolete.

SELECTED BIBLIOGRAPHY

Books and Monographs

Barendt, Eric. Freedom of Speech. 2d ed. Oxford University Press, 2005.

Collins, Matthew. The Law of Defamation and the Internet. 3d ed. Oxford University Press, 2010. Edwards, Lilian, and Charlotte Waelde. Law and the Internet. 3d ed. Hart Publishing, 2009.

Pasquale, Frank. The Black Box Society: The Secret Algorithms That Control Money and Information.

Harvard University Press, 2015.

Solove, Daniel J. The Future of Reputation: Gossip, Rumor, and Privacy on the Internet. Yale University Press, 2007.

Journal Articles

Ardia, David S. Reputation in a Networked World: Revisiting the Social Foundations of Defamation Law.

45 Harv. C.R.-C.L. L. Rev. 261 (2010).

Lidsky, Lyrissa Barnett. Silencing John Doe: Defamation and Discourse in Cyberspace. 49 Duke L.J. 855 (2000).

Rosen, Jeffrey. The Right to Be Forgotten. 64 Stan. L. Rev. Online 88 (2012). Sylvain, Olivier. Intermediary Design Duties. 50 Conn. L. Rev. 203 (2018).

Cases

Communications Decency Act § 230, 47 U.S.C. § 230 (2018).

Defamation Act 2013, c. 26 (Eng.).

Delfi AS v. Estonia, App. No. 64569/09 (Eur. Ct. H.R. June 16, 2015). Dow Jones & Co. v. Gutnick (2002) 210 CLR 575 (Austl.).

Gertz v. Robert Welch, Inc., 418 U.S. 323 (1974).

Google Inc. v. Duffy (2017) 129 SASR 304 (Austl.).

Google LLC v. Equustek Solutions Inc., 2017 SCC 34 (Can.).

Google Spain SL v. Agencia Española de Protección de Datos, Case C-131/12, ECLI:EU:C:2014:317 (May 13, 2014).

Lee Hsien Loong v. Leong Sze Hian [2020] SGHC 208 (Sing.). Monroe v. Hopkins [2017] EWHC 433 (QB) (Eng.).

New York Times Co. v. Sullivan, 376 U.S. 254 (1964).

Rachel Ehrenfeld v. Mahfouz, No. 04 Civ. 9641, 2006 WL 1096816 (S.D.N.Y. 2006). Reynolds v. Times Newspapers Ltd [1999] UKHL 45 (Eng.).

Stocker v. Stocker [2019] UKSC 17 (Eng.).

Subramanian Swamy v. Union of India, (2016) 7 SCC 221 (India).

Yahoo! Inc. v. La Ligue Contre le Racisme et l'Antisémitisme, 433 F.3d 1199 (9th Cir. 2006).

International Instruments

International Covenant on Civil and Political Rights, Dec. 16, 1966, 999 U.N.T.S. 171. Regulation (EU) 2016/679 (General Data Protection Regulation), 2016 O.J. (L 119) 1. Council Regulation 2022/2065 (Digital Services Act), 2022 O.J. (L 277) 1.

SPEECH Act, 28 U.S.C. §§ 4101–4105 (2010).

[1]Eric Barendt, Freedom of Speech 1–12 (2d ed. 2005).

[2]Matthew Collins, The Law of Defamation and the Internet 1 (3d ed. 2010).

3Eric Barendt, Freedom of Speech 1–12 (2d ed. 2005).

[4]New York Times Co. v. Sullivan, 376 U.S. 254, 279–80 (1964).

5 Matthew Collins, The Law of Defamation and the Internet 1 (3d ed. 2010).

[6]Dow Jones & Co. v. Gutnick (2002) 210 CLR 575, 600 (Austl.).

[7]Daniel J. Solove, The Future of Reputation: Gossip, Rumor, and Privacy on the Internet 78 (2007).

[8]See generally David S. Ardia, Reputation in a Networked World: Revisiting the Social Foundations of Defamation Law, 45 Harv. C.R.-C.L. L. Rev. 261 (2010).

[9]Olivier Sylvain, Intermediary Design Duties, 50 Conn. L. Rev. 203, 214 (2018). [10]Council Regulation 2022/2065, art. 16, 2022 O.J. (L 277) 1 (EU) (Digital Services Act). [11]Lilian Edwards & Charlotte Waelde, Law and the Internet 342 (3d ed. 2009).

[12]Lyrissa Barnett Lidsky, Silencing John Doe: Defamation and Discourse in Cyberspace, 49 Duke L.J. 855, 858 (2000).

[13]Jeffrey Rosen, The Right to Be Forgotten, 64 Stan. L. Rev. Online 88, 89 (2012).

[14]Regulation (EU) 2016/679, art. 17, 2016 O.J. (L 119) 1 (GDPR) (right to erasure).

[15]Frank Pasquale, The Black Box Society: The Secret Algorithms That Control Money and Information 59–60 (2015).

[16]Restatement (Second) of Torts § 558 (Am. L. Inst. 1977).

[18]Monroe v. Hopkins [2017] EWHC 433 (QB) (Eng.).

[20]Stocker v. Stocker [2019] UKSC 17, [2019] AC 650 (appeal taken from Eng.).

[21]Defamation Act 2013, c. 26, § 1(1) (Eng.).

[22]Subramanian Swamy v. Union of India, (2016) 7 SCC 221 (India).

[23]Int'l Covenant on Civil and Political Rights arts. 17, 19, Dec. 16, 1966, 999 U.N.T.S. 171.

[24]Delfi AS v. Estonia, App. No. 64569/09, 1–10 (Eur. Ct. H.R. June 16, 2015).

[25]Lee Hsien Loong v. Leong Sze Hian [2020] SGHC 208 (Sing.).

[28]Communications Decency Act § 230, 47 U.S.C. § 230 (2018).

[30]Information Technology Act, 2000, § 66A, No. 21 of 2000 (India) (provision struck down); Indian Penal Code, 1860, §§ 499–500 (India).

[32]Google LLC v. Equustek Solutions Inc., 2017 SCC 34, [2017] 1 S.C.R. 824 (Can.).

[36]Google Spain SL v. Agencia Española de Protección de Datos, Case C-131/12, ECLI:EU:C:2014:317, ¶ 88 (May 13, 2014).

[37]Reynolds v. Times Newspapers Ltd [1999] UKHL 45, [2001] 2 AC 127 (appeal taken from Eng.).

[38]Gertz v. Robert Welch, Inc., 418 U.S. 323, 347 (1974).

[40]Google Inc. v. Duffy (2017) 129 SASR 304 (Austl.).

41 Dow Jones & Co. v. Gutnick (2002) 210 CLR 575 (Austl.).

42 Regulation (EU) 2016/679 (General Data Protection Regulation), 2016 O.J. (L 119) 1

[43]Rachel Ehrenfeld v. Mahfouz, No. 04 Civ. 9641, 2006 WL 1096816 (S.D.N.Y. Apr. 26, 2006); see also SPEECH Act, 28 U.S.C. §§ 4101–4105 (2010).

[45]Yahoo! Inc. v. La Ligue Contre le Racisme et l'Antisémitisme, 433 F.3d 1199 (9th Cir. 2006).

47 Lee Hsien Loong v. Leong Sze Hian [2020] SGHC 208 (Sing.).

48 Council Regulation 2022/2065 (Digital Services Act), 2022 O.J. (L 277) 1.

49 International Covenant on Civil and Political Rights, Dec. 16, 1966, 999 U.N.T.S. 171.