PIONEER
ACADEMIC PUBLISHING LIMITED

Studies in **Law and Justice**

Volume 4, Number 5 **October 2025**

CONTENTS

PIONEER
ACADEMIC PUBLISHING LIMITED

Studies in Law and Justice
ISSN 2958-0382
www.pioneerpublisher.com/slj
Volume 4 Number 5 October 2025

Analysis of *Breen v Williams*: A Critical Examination of the Doctor-Patient Fiduciary Relationship

Junlin Wang[1]

[1] LL.M., The University of Sydney, Camperdown NSW 2050, Australia
Correspondence: Junlin Wang, LL.M., The University of Sydney, Camperdown NSW 2050, Australia.

doi:10.56397/SLJ.2025.10.01

Abstract

This paper critically examines *Breen v Williams*, where the High Court of Australia rejected the view that doctors owe fiduciary duties to grant patients access to medical records. By restricting fiduciary obligations to negative duties, the Court overlooked patient vulnerability, trust, and the therapeutic realities of healthcare. Drawing on Canadian jurisprudence, particularly *McInerney v MacDonald*, the paper highlights the divergence from international approaches that affirm positive disclosure duties. It argues for a reformed fiduciary framework aligned with patient autonomy, while noting that legislative reforms, such as the Health Records and Information Privacy Act 2002 (NSW), have partially addressed these gaps.

Keywords: *Breen v Williams*, doctor-patient relationship, fiduciary duty, access to medical records, patient rights, information privacy

1. Introduction

The landmark case of *Breen v Williams* represents a vital moment in Australian medical law with regard to the nature of the doctor-patient relationship. However, the decision has been criticised for failing to acknowledge the full complexity of the doctor-patient relationship. In this case, the High Court unanimously rejected the argument that a doctor undertakes a fiduciary duty to grant patients access to their medical records. While this approach may reflect a traditional understanding of fiduciary law, it fails to capture the relational, informational, and moral dimensions that define modern healthcare.

In light of this criticism, this paper analyses the High Court's reasoning in *Breen*, highlighting its doctrinal limitations and practical consequences. It examines relevant case law and academic commentary, including Canadian jurisprudence and post-*Breen* Australian developments. It also considers reform proposals such as the Appointed Fiduciary Model and Fiduciary Informed Consent. This paper agrees that a reformed, context-sensitive fiduciary framework is essential to better protect patient rights and reflect the relational realities of modern healthcare.

2. The High Court's Reasoning

In *Breen v Williams*, Ms Breen was a former patient of Dr. Williams. She requested permission to review and copy all the medical

records of her diagnosis and treatment for another class action in the United States. The High Court rejected her claim based on five grounds, including that no fiduciary duty obliged Dr. Williams to provide such access.

The majority, particularly Brennan CJ and Gaudron and McHugh JJ, held that the fiduciary relationship between doctors and patients occurs in limited circumstances. Brennan CJ stated that fiduciary duties do not impose positive obligations to act in another's interests, but rather restrain a fiduciary from acting against the interests of the principal.[1] Gaudron and McHugh JJ asserted that where the doctor-patient relationship is characterised by trust, confidence and the patient's inherent vulnerability, fiduciary obligations arise.[2] It mainly relates to the doctor's conduct in providing diagnosis, advice, and treatment.[3] Although the judges acknowledged aspects of trust and vulnerability in the doctor-patient relationship in this case, they found that these were insufficient to ground a fiduciary duty of disclosure.[4] What's more, they also mentioned the Canadian case of *McInerney v MacDonald*, where a fiduciary duty to disclose medical records was recognised. However, the High Court distinguished it on legal grounds and ultimately declined to follow its reasoning.[5]

Furthermore, the High Court also distinguished between proscriptive (negative) and prescriptive (positive) duties within the doctor-patient relationship. The court held that doctors owe proscriptive fiduciary duties such as the duty to avoid conflicts of interest and to maintain patient confidentiality.[6] But the Court firmly rejected the notion that fiduciary nature of the doctor-patient relationship gives rise to prescriptive duties. In particular, it held that there is no enforceable obligation requiring doctors to grant patients access to their medical records. Brennan CJ stated that while a doctor may be required to provide medical information in certain circumstances—such as where refusal may prejudice the patient's health—this obligation does not extend to granting direct access to or copies of medical records.[7] Arguably, the Court's reasoning affirms that fiduciary duties in the medical context are essentially proscriptive, and any prescriptive duties must be established through legislative reform rather than judicial implication.[8]

3. Criticisms

3.1 Narrow View of Medical Relationships

Legal scholars have argued that the ruling of *Breen v Williams* reflects a misunderstanding of the doctor-patient relationship in the therapeutic reality. Zara J Bending asserts that the High Court's adherence to a narrow contractual model ignores the significance of trust in this relationship.[9] She points out an insufficient recognition of power imbalance in this case. While Brennan CT acknowledged that 'the doctor-patient relationship is one in which the doctor has an ascendancy over the patient and the patient places trust in the doctor,'[10] the court failed to translate this recognition into a comprehensive legal acknowledgment of the doctor-patient relationship's nature.[11]

Some legal scholars also criticised this ruling as overly formalistic. Diana Nestorovska argues that the High Court's rigid application of fiduciary categories ignored the reality that patients entrust their wellbeing, bodies, and personal information to physicians.[12] She observes that although fiduciary law is intended to protect against the abuse of power and dependence, the Court's refusal to recognise any affirmative obligations in *Breen* ultimately undermines equity's protective purpose.[13] She also notes that the judgment reinforces a view of fiduciary obligations as purely proscriptive, which differs from modern ethical and relational approaches to healthcare.[14]

In contrast, the Canadian Supreme Court has taken a more expansive view. In *McInerney v*

[1] *Breen v Williams* (1996) 186 CLR 71, 83 (Brennan CJ).

[2] Ibid 107 (Gaudron and McHugh JJ).

[3] Ibid 107 (Gaudron and McHugh JJ).

[4] Ibid 108 (Gaudron and McHugh JJ).

[5] *McInerney v MacDonald* [1992] 2 SCR 138, 148-9.

[6] *Breen v Williams* (1996) 186 CLR 71, 83 (Brennan CJ), 92-93 (Dawson and Toohey JJ).

[7] Ibid 79 (Brennan CJ).

[8] Ibid 83 (Brennan CJ), 92–94 (Dawson and Toohey JJ).

[9] Zara J Bending. (2015). Reconceptualising the Doctor-Patient Relationship: Recognising the Role of Trust in Contemporary Health Care. *Journal of Bioethical Inquiry*, *12*(2), 189, 190.

[10] *Breen v Williams* (1996) 186 CLR 71, 83 (Breena CJ).

[11] Bending (n 9) 190.

[12] Diana Nestorovska. (2018). Revisiting Breen v Williams: Breathing Life into a Doctor-Patient Fiduciary Relationship. *Journal of Law and Medicine*, *25*(4), 692, 695-6.

[13] Ibid 696.

[14] Ibid 695.

MacDonald, La Forest J held that while medical records may be physically owned by the physician, their contents belong to the patient, and must be disclosed unless disclosure would be harmful.[1] This contrasts starkly with the High Court's property-based rationale in *Breen*. Moreover, in *Norberg v Wynrib*, the Canadian Court found that fiduciary duties applied to a doctor who took advantage of a patient's drug dependency. Therefore, the Court reasoned that the trust and power imbalance in the medical relationship created obligations that went beyond those found in contract or tort law.[2]

3.2 Overly Rigid Distinction of Proscriptive and Prescriptive

The High Court's commitment to the proscriptive/prescriptive distinction has been challenged. In *Breen v Williams*, the Court held that fiduciary duties are limited to proscriptive obligations, obliging the fiduciary only to refrain from obtaining unauthorised benefits or acting in conflict with the interests of the beneficiary.[3] This rigid approach has been criticized for being unrealistic and inadequate in situations where protecting vulnerable individuals calls for affirmative action.[4] Justice Fabian Gleeson points out that although fiduciary duties are said to be purely proscriptive in Australia, courts have often recognised positive obligations (such as disclosure, candour, and loyalty).[5] Importantly, this recognition is not limited to interpersonal contexts; it also extends to duties owed by corporate fiduciaries. In such cases, courts have enforced these obligations even where they require affirmative conduct—for example, taking steps to avoid conflicts of interest through proactive disclosure.[6]

Furthermore, the doctrinal tension is evident in decisions such as *Duncan v Independent Commission Against Corruption*, Justice McDougall acknowledged that proactive measures may be required to fulfil fiduciary duty to avoid conflicts.[7] The *Duncan* expose the limitations of a strict proscriptive model and suggests that functional needs, rather than rigid legal formalism, should guide the identification of fiduciary duties.

3.3 Failure to Recongnise the Therapeutic Reality

Beyond its rigid legal framework, some scholars have challenged *Breen* for failing to reflect the substantive realities faced by patients. Diana Nestorovska argues that the Court obscures the inherently unequal and trust-based nature of medical care.[8] She also emphasises that equity is designed to respond to such vulnerability, and its retreat in *Breen* represents a failure to protect patients' right in the face of informational imbalance.[9]

Zara J Bending similarly criticises the Court for relying on a commercial contract model, which treats the doctor-patient relationship as a transaction between equals.[10] In doing so, the High Court imposed an inappropriate "market logic" that does not suit the healthcare context.[11] By comparison, Bending proposes a "Trust Model" that acknowledges the ethical and relational dimensions of medical care. She argues that patients expect not merely technical competence but also good faith, disclosure of conflicts, and the prioritisation of their interests.[12]

3.4 Inconsistency with International Approach

In addition to domestic scholarly criticism, the High Court's reasoning in *Breen v Williams* places Australian fiduciary jurisprudence at odds with the evolving international trends that favour a more expansive and protective conception of the doctor-patient relationship. In *McInerney v MacDonald*, it demonstrates that Canadian courts have increasingly embraced a functional and ethically grounded fiduciary model. In particular, doctors are under affirmative duties to disclose, inform, and prioritise their patients' interest. As mentioned above, the Supreme Court of Canada held that doctors may retain physical ownership of medical records. However, the informational

[1] *McInerney v MacDonald* [1992] 2 SCR 138, 148-9 (La Forest J).

[2] *Norberg v Wynrib* [1992] 2 SCR 226, 272 (McLachlin J).

[3] *Breen v Williams* (1996) 186 CLR 71, 113 (Gaudron and McHugh JJ).

[4] The Honourable Justice Fabian Gleeson. (2018). Proscriptive and Prescriptive Duties: Is the Distinction Helpful and Sustainable, and If So, What Are the Practical Consequences? *Judicial Review*, *14*, 70, 73-4.

[5] Ibid 74.

[6] Ibid 75-6.

[7] *Duncan v Independent Commission Against Corruption* [2014] NSWSC 1018, [205] (McDougall J).

[8] Nestorovska (n 12) 694-6.

[9] Ibid 697.

[10] Bending (n 11) 190-1.

[11] Ibid 193.

[12] Ibid 194-6.

content belongs to the patient, giving rise to a fiduciary duty of disclosure.[1] La Forest J observed that "the trust reposed in the physician by the patient mandates that the flow of information operate both ways" and that disclosure of medical records is essential to preserving that trust.[2] The judgment clearly rejected the idea that fiduciary duties must be passive or solely restrictive. Instead, it viewed access to information as a fundamental aspect of fiduciary loyalty and a key element in respecting patient autonomy.

By contrast, the High Court in *Breen* declined to adopt these protections, dismissing *McInerney* as inconsistent with Australian law.[3] This divergence raises broader questions about Australia's intolerance in fiduciary reasoning. The aim of fiduciary law is to protect individuals who are vulnerable to the discretion of others. Australia's rejection of prescriptive duties therefore risks undermining the fundamental purpose of fiduciary protection in the medical context.

4. Doctrinal and Practical Consequences

4.1 Inadequate Protection of Patient Rights

As a result of the restrictive approach adopted in *Breen v Williams*, a doctrinal gap has emerged in the protection of patient rights. By refusing to recognise affirmative fiduciary duties, the decision offers no legal mechanism. Arguably, there is no legal mechanism to ensure that doctors provide patients with access to information, support their decision-making, or prioritise their best interests. As Diana Nestorovska observes, the Court's narrow view of fiduciary law overlooks the trust, dependency, and informational asymmetry that define modern health care.[4] Patients entrust their well-being, bodies, and confidential information to doctors—yet *Breen* provides no positive obligations to match that trust.[5]

4.2 Regressive Effect on the Doctor-Patient Relationship

The ruling also risks undermining the development of ethical and relational models of care. Zara J Bending critiques the Court for applying a "market logic" to the therapeutic context, treating the relationship as a transaction between equals.[6] This framing ignores the emotional, ethical, and informational dimensions of medical care, where patients often seek support, reassurance, and shared decision-making. Without a legal duty to disclose or guide, the doctor-patient relationship may default to formality and detachment, leaving patients without the relational support they need.[7]

4.3 Doctrinal Rigidity and Erosion of Fiduciary Law

Finally, *Breen* imposes a rigid proscriptive model on fiduciary law that undermines its broader protective purpose. Fiduciary doctrine is designed to evolve across contexts, protecting beneficiaries wherever there is vulnerability and power imbalance.[8] Yet by declaring fiduciary duties to be inherently proscriptive, the High Court restricts this flexibility and risks causing fiduciary law to become rigid and outdated.[9] Moreover, Gleeson J has criticised *Breen*'s formulation as being more formalistic than substantive. He points out that in many real-world scenarios, a failure to act—such as a failure to inform or guide—may have the same practical consequences as disloyal or self-interested conduct.[10] He argues that a rule which denies fiduciary liability for omissions, even when they amount to betrayal of trust, undermines the very purpose of fiduciary law.

By locking fiduciary doctrine into a formalistic and static framework, *Breen* weakens its capacity to respond to contemporary ethical and relational challenges. A failure to update this approach may erode public confidence in fiduciary protections, and diminish the law's normative force in contexts where it is most needed.

5. Recommendations for Reform

In light of the doctrinal shortcomings exposed by *Breen v Williams*, several models of reform should be considered to protect patients' rights. These reforms aim to re-align fiduciary law in medical settings to recognise patients' inherent dependence and vulnerability. Besides, they also

[1] *McInerney v MacDonald* [1992] 2 SCR 138, 148-9.

[2] Ibid 149 (La Forest J).

[3] *Breen v Williams* (1996) 186 CLR 71, 112-13 (Gaudron and McHugh JJ).

[4] Nestorovska (n 12) 696-7.

[5] Ibid 695-6.

[6] Bending (n 9) 193-4.

[7] Ibid 194-6.

[8] *Hospital Products Ltd v United States Surgical Corporation* (1984) 156 CLR 41, 97 (Mason J).

[9] *Breen v Williams* (1996) 186 CLR 71, 113 (Gaudron and McHugh JJ); see also Gleeson (n 18) 74.

[10] Gleeson (n 18) 75-6.

acknowledge that doctors possess both the specialised expertise and the ethical duty to guide patients in making informed treatment decisions.

The first one is the appointed fiduciary model developed by Davies and Parker. In their opinions, patients can choose to formally allow their doctor to help make decisions on their behalf.[1] This approach does not assume that every doctor-patient relationship works the same way. Instead, the doctor may act mainly as an adviser, or may take a more active role if the patient requests it.[2] This model agrees that not all patients want to make decisions alone especially those feel overwhelmed or uncertain. Moreover, it provides a respectful and structured way for doctors to help without taking away the patient's freedom to decide.

Another recommendation comes from Ludewig and colleagues. They propose a more ethical and patient-centred approach to medical decision-making. They suggest that the doctor-patient relationship should be based on mutual trust and shared responsibility, not just on legal duties.[3] In consequence, they introduce the idea of fiduciary informed consent. Under this model, doctors should not only provide information but also take time to understand the patient's values, needs and preferences. What's more, doctors should take the responsibility to help the patients make decision that reflect what matters most of them.[4] This does not mean doctors should make decisions for patients, but rather that they should work with patients to guide them in a respectful and supportive way.

Both models challenge the rigid legal rule in *Breen*, which says doctors only need to avoid conflicts of interest or personal gain. Instead, it is suggested that doctors sometimes have a positive duty to assist, inform, and guide, especially when patients ask for help. This would bring fiduciary law closer to modern expectations in healthcare, where patients are treated as partners, not just as recipients of care.

[1] Ben Davies and Joshua Parker. (2022). Doctors as Appointed Fiduciaries: A Supplemental Model for Medical Decision-Making. *Cambridge Quarterly of Healthcare Ethics*, *31*(1), 23, 24.

[2] Ibid 25-6.

[3] Sophie Ludewigs et al. (2025). Ethics of the Fiduciary Relationship Between Patient and Physician: The Case of Informed Consent. *Journal of Medical Ethics*, *51*(1), 59, 60-1.

[4] Ibid 62-3.

6. Legislative and Policy Development

6.1 Legislation

After the High Court declined to recognise a fiduciary right of access to medical records in *Breen v Williams*, legislative measures soon became the dominant approach to advancing reform. In New South Wales, the *Health Records and Information Privacy Act 2002* (NSW) grants patients a legally enforceable right to access their personal health information. These include situations where access would pose a serious threat to life or health, unreasonably impact the privacy of others, prejudice legal proceedings or investigations, reveal sensitive negotiation intentions, or where access is prohibited or required to be denied by law.[5] Similarly, the *Privacy Act 1988* (Cth), through the *Australian Privacy Principles* (APPs), provides comprehensive access and correction rights about health information held by Commonwealth-regulated entities.[6] All of these provide strong safeguards for protecting the legitimate rights and interests of patients in the medical-patient relationship. These statutory developments have filled the normative gap left by *Breen*, establishing patient access rights as a matter of privacy and administrative fairness rather than equity.

6.2 Case

Subsequent case law demonstrates a judicial trend toward strengthening the obligations owed by medical professionals to patients. It is with regret that this evolution has primarily occurred within the boundaries of negligence law, rather than through a direct expansion of fiduciary duties as recognised at common law.

In *Cattanach v Melchior*, the Court examined a situation where a doctor failed to inform a patient about an undiscovered fallopian tube before a sterilisation procedure.[7] The judges considered the issue whether the doctor could be held liable for the financial costs of raising a healthy child as a result. The majority held that a doctor could be found liable in negligence for failing to provide material information that might affect a patient's reproductive choices. While the judgment underscored the importance of disclosure in sensitive medical contexts, the

[5] *Health Records and Information Privacy Act 2002* (NSW) pt 4, div 1.

[6] *Privacy Act 1988* (Cth) sch 1.

[7] *Cattanach v Melchior* (2003) 215 CLR 1, 41-42 (Gummow and Kirby JJ), 95 (Callinan J).

Court did not identified this duty as fiduciary in nature. Instead, the obligation was based on the tortious duty of care. This principle requires doctors to exercise reasonable skill and to disclose information that a reasonable person in the patient's position would want to know.[1]

In *Hunter Area Health Service v Presland*, the NSW Court of Appeal rejected a psychiatric patient's claim in negligence after he committed a violent act following hospital discharge.[2] The Court declined to impose a duty of care in the circumstances and made no attempt to frame the medical professionals' obligations as fiduciary. Rather than extending fiduciary doctrines, the case shows that judges are reluctant to broaden liability in mental health situations. This is especially apparent in cases involving complex clinical decisions and matters of public interest.

By contrast, *Rogers v Whitaker* marked a significant shift in how courts view doctors' disclosure obligations. The High Court held that a doctor's failure to warn a patient of a rare risk of blindness amounted to negligence, even where the risk was less than one percent.[3] The Court affirmed that patients have the right to make informed decisions about their medical treatment. It also held that a doctor's duty to warn is not governed by medical custom, but by whether the information is materially significant to the individual patient.[4] Although the case was decided in negligence, scholars have interpreted it as laying the foundation for a more relational and patient-centred approach. This interpretation aligns with the underlying rationale of fiduciary obligations.[5]

Thus, while Australian courts have been reluctant to formally expand fiduciary duties in the medical context beyond their traditional proscriptive scope, cases such as *Rogers v Whitaker* reflect a growing judicial emphasis on affirmative duties. This shift aligns with ethical principles that support patient autonomy, trust, and recognition of vulnerability.

7. Conclusion

The High Court's decision in *Breen v Williams* marks a pivotal moment in Australian medical law. However, its restrictive interpretation of fiduciary obligations in the doctor-patient relationship remains deeply problematic. By maintaining a rigid distinction between proscriptive and prescriptive duties, the Court overlooked the nuanced realities of patient vulnerability and the dynamics of trust inherent in contemporary healthcare. This has left significant gaps in legal protection for patients, particularly where breaches of trust do not involve personal interests.

This position diverges sharply from international jurisprudence. Canada Courts there have recognised that fiduciary duties in healthcare may require affirmative action, such as disclosure, guidance, and prioritisation of patient interests. While subsequent Australian case law has shown a growing judicial appreciation of patient autonomy and trust, this has occurred mainly in areas such as informed consent and medical disclosure. What's more, this evolution remains limited in its reach. These developments have generally taken place within the boundaries of negligence law, rather than through a direct expansion of fiduciary doctrine.

The ongoing reliance on legislative intervention, such as the *Health Records and Information Privacy Act 2002* (NSW) and the *Privacy Act 1988* (Cth), has begun to address the gaps left by *Breen*. However, statutory reforms alone cannot resolve the doctrinal and relational shortcomings of the High Court's approach. Reform models—such as the Appointed Fiduciary framework and Fiduciary Informed Consent—provide a blueprint for recalibrating fiduciary law to better reflect patient dependence, autonomy, and the ethical responsibilities of doctors.

In sum, a broader, context-sensitive reframing of fiduciary duties in the medical sphere is both possible and necessary. Fiduciary law can fulfil its purpose only by moving beyond formalistic limitations. It would benefit from adopting a more relational and protective approach. This approach ensures that people who place their trust and wellbeing in the hands of medical professionals are meaningfully protected in both law and practice.

References

Ben Davies and Joshua Parker. (2022). Doctors as Appointed Fiduciaries: A Supplemental Model for Medical Decision-Making.

[1] Ibid 16-17 (Gleeson CJ).

[2] *Hunter Area Health Service v Presland* [2005] NSWCA 33, [100]-[110] (Spigelman CJ); see also [21]-[22], [142]-[144] (Ipp JA).

[3] *Rogers v Whitaker* (1992) 175 CLR 479, 489-490 (Brennan J).

[4] Ibid 490-491.

[5] McLean, Sheila. (2002). What Do We Mean by "Fiduciary Duty" in Medical Law? *Medical Law Review*, *20*(2), 129, 136-138.

Cambridge Quarterly of Healthcare Ethics, *31*(1), 23, 24.

Diana Nestorovska. (2018). Revisiting Breen v Williams: Breathing Life into a Doctor-Patient Fiduciary Relationship. *Journal of Law and Medicine*, *25*(4), 692, 695-6.

McLean, Sheila. (2002). What Do We Mean by "Fiduciary Duty" in Medical Law? *Medical Law Review*, *20*(2), 129, 136-138.

Sophie Ludewigs et al. (2025). Ethics of the Fiduciary Relationship Between Patient and Physician: The Case of Informed Consent. *Journal of Medical Ethics*, *51*(1), 59, 60-1.

The Honourable Justice Fabian Gleeson. (2018). Proscriptive and Prescriptive Duties: Is the Distinction Helpful and Sustainable, and If So, What Are the Practical Consequences? *Judicial Review*, *14*, 70, 73-4.

Zara J Bending. (2015). Reconceptualising the Doctor-Patient Relationship: Recognising the Role of Trust in Contemporary Health Care. *Journal of Bioethical Inquiry*, *12*(2), 189, 190.

PIONEER
ACADEMIC PUBLISHING LIMITED

Studies in Law and Justice
ISSN 2958-0382
www.pioneerpublisher.com/slj
Volume 4 Number 5 October 2025

The "Intertwined" Relationship and Deadlock Between Delimitation of the Continental Shelf Beyond 200 nm and the Delineation of Its Outer Limits

Zichun Tang[1]

[1] Wuhan University, Wuhan 430000, China
Correspondence: Zichun Tang, Wuhan University, Wuhan 430000, China.

doi:10.56397/SLJ.2025.10.02

Abstract

This article examines the legal relationship and practical challenges between "delimitation" and "delineation" within the framework of the United Nations Convention on the law of the sea. Through an analysis of rulings by the ITLOS and the contrasting stance adopted by the ICJ, it points out that international judicial bodies have yet to clarify the relationship between the two. Although Articles 76(10) and 83(1) of the UNCLOS treat the two as independent processes in form, in practice they are deeply intertwined due to overlapping entitlements, geographical foundations, and procedural interactions. To break the deadlock—where coastal States are trapped in a cycle of being unable to achieve successful delimitation without first delineating outer limits, a situation exacerbated by Paragraph 5(a) of Annex I to the Rules of Procedure of the Commission on the Limits of the Continental Shelf—this article assesses the interim solution proposed by scholars, and ultimately proposes a fundamental solution: reforming Paragraph 5(a) by recommending that the CLCS be empowered to conduct preliminary technical considerations of all submissions.

Keywords: delimitation, delineation, the Commission on the Limits of the Continental Shelf (CLCS), United Nations Convention on the Law of the Sea (UNCLOS), Paragraph 5(a) of Annex I to the Rules of Procedure of CLCS

1. Introduction

Since International Tribunal for the Law of the Sea (ITLOS) affirmed its jurisdiction over delimitation of the continental shelf beyond 200 nautical miles (nm) in the *Delimitation of the maritime boundary in the Bay of Bengal* (*Bangladesh/Myanmar Case*), the relationship between "delimitation of outer continental shelf" and "delineation" has drawn increasing attention. Yet the Tribunal did not fully settle the matter. While it agreed to exercise jurisdiction, it did so under strict conditions, without clearly defining the relationship or sequence between the two processes. Rather, it issued a situation-specific ruling, which has since drawn criticism.

Why does the relationship between "delimitation" and "delineation" matter?

Because clarifying it is key to resolving the current impasse—where delimitation and the delineation often work at cross purposes. Under Paragraph 5(a) of Annex I to the Rules of Procedure of the Commission on the Limits of the Continental Shelf,[1] CLCS may not consider or qualify a submission in cases where a land or maritime dispute exists.[2] This means that if any of the States concerned in the dispute objects, the coastal state's submission is stalled, leaving it unable to obtain final and binding outer limits. If the delineation is a prerequisite for delimitation, the coastal state cannot proceed with maritime boundary delimitation. If it is not, the state could theoretically delimit without going through the CLCS. The underlying concern is: if other states concerned can freely block the CLCS process, can they also arbitrarily prevent delimitation of the continental shelf?

In practice, however, even when an international court or tribunal delimits the continental shelf first, the deadlock may remain. For instance, Myanmar made its submission to the CLCS in 2008. Afterwards, Bangladesh objected in 2009. In 2012, ITLOS rendered its judgment, resolving the delimitation dispute between the two countries. Yet that same year, Bangladesh still refused to consent to CLCS consideration of Myanmar's submission, arguing that the judgment had not fully settled the outer limits.[3] However, the reason the judgment only determined one segment and the direction of the boundary line was precisely that, without the recommendations from the CLCS on the outer limits of the continental shelf, the Tribunal could not determine the other endpoint of the boundary line. This perfectly illustrates the Catch-22.

2. Answers in the Cases

In the *Bangladesh/Myanmar Case*, the Tribunal concluded that exercising jurisdiction to delimit the continental shelf beyond 200 nm first would not prejudice the subsequent delineation of outer limits by the CLCS, reasoning as follows:

First, there is a clear distinction between the delimitation of the continental shelf under article 83 and the delineation of its outer limits under article 76.[4] The absence of established outer limits of a maritime zone does not preclude delimitation of that zone.[5] And the exercise by international courts and tribunals of their jurisdiction regarding the delimitation of maritime boundaries is without prejudice to the exercise by the CLCS of its functions on matters related to the delineation.[6]

Second, there is nothing in United Nations Convention on the Law of the Sea (UNCLOS) or in the Rules of Procedure or in its practice to indicate that delimitation of the continental shelf constitutes an impediment to the performance by the CLCS of its functions.[7]

Third, a number of States have resolved their maritime boundaries through negotiation or agreements either before or during the CLCS's consideration of their submissions, and such practice has not precluded the CLCS's examination or issuing recommendations.[8]

Fourth, if a court or tribunal were to decline jurisdiction over delimitation, while the disputing States simultaneously withhold consent for CLCS consideration of a submission, the dispute could be prolonged indefinitely, leaving the Parties in a position where they may be unable to benefit fully from their rights over the continental shelf.[9]

Similarly, in *Matter of the Bay of Bengal Maritime Boundary Arbitration* (*Bangladesh/India Case*), the arbitral tribunal emphasized that "if the Tribunal were to decline to delimit the continental shelf beyond 200 nm, the outer limits of the continental shelf of each of the Parties would remain unresolved."[10]

Although ITLOS prioritized delimitation over the delineation on outer limits in the

[1] In the following text, Paragraph 5(a) of Annex I to the Rules of Procedure of the Commission on the Limits of the Continental Shelf is abbreviated as Paragraph 5(a); the Rules of Procedure of the Commission on the Limits of the Continental Shelf is abbreviated as the Rules of Procedure; the Commission on the Limits of the Continental Shelf is abbreviated as CLCS.

[2] Article 5 paragraph a of Annex I, the Rules of Procedure of the Commission on the Limits of the Continental Shelf.

[3] Note Verbale from the Permanent Mission of Bangladesh to the United Nations, No. PMBNY/67/UNCLOS/2012, 30 September 2012, https://www.un.org/depts/los/clcs_new/submissions_files/mmr08/2012_09_30_BGD_NV_UN.pdf.

[4] Delimitation of the maritime boundary in the Bay of Bengal (Bangladesh/Myanmar), Judgment, ITLOS Reports 2012, at 99, para. 376.

[5] Ibid, 98, para. 370.

[6] Ibid, 100, para. 379.

[7] Ibid, 99, para. 377.

[8] Ibid, 100, 102, para. 380, 393.

[9] Ibid, 102, para. 392.

[10] In the Matter of the Bay of Bengal Maritime Boundary Arbitration (Bangladesh/India), Award, 7 July 2014, at 22, para. 82.

Bangladesh/Myanmar Case, the two cases regarding Bay of Bengal did not offer a definitive or universal solution. The ITLOS Tribunal also acknowledged that where significant uncertainty exists regarding the outer edge of the continental margin, it would hesitate to proceed with delimitation.[1]

The approach of the International Court of Justice (ICJ), however, marks a reversal from the reasoning adopted in the two cases regarding Bay of Bengal. In *Question of the Delimitation of the Continental Shelf between Nicaragua and Colombia beyond 200 Nautical Miles from the Nicaraguan Coast* (*Nicaraguan Coast Case*), the ICJ interpreted its 2012 judgment regarding *Territorial and Maritime Dispute (Nicaragua v. Colombia)* as premising the delimitation of the continental shelf beyond 200 nautical miles on Nicaragua's submission of data on its outer continental shelf limits to the Commission. [2]Judge Xue Hanqin, in her separate opinion on *Nicaraguan Coast Case*, also stated: "In such a technically complicated case, it is a necessity for the parties to obtain the recommendations of the CLCS before proceeding to delimitation." [3] Clearly, in their view, delimitation of the outer continental shelf should be premised on the delineation of its outer limits, since the CLCS, as the authoritative scientific and technical body, can provide geographical and legal basis for judicial organs through its consideration and recommendations.

3. "Delimitation" and "Delineation" in UNCLOS

According to Article 76(10) of UNCLOS, the delineation of the outer limits of the continental shelf "is without prejudice to the question of delimitation of the continental shelf between States with opposite or adjacent coasts." Under Article 83(1), delimitation of the continental shelf between States with opposite or adjacent coasts "shall be effected by agreement on the basis of international law, as referred to in Article 38 of the Statute of the ICJ, in order to achieve an equitable solution." Based on these two provisions, "delimitation" and "delineation" are two independent processes.

However, such a straightforward conclusion overlooks the evolving and complex interplay between the two processes. According to the *travaux préparatoires*, the United States first proposed the establishment of the Commission during the Third United Nations Conference on the Law of the Sea. In the U.S. proposal, the Commission was envisaged as an institution that would review and finally decide on the outer limits, exercising a function similar to that of a court rendering a judgment.[4] However, as the drafting progressed, the powers granted to the Commission were gradually scaled back. The final text of the UNCLOS only requires the coastal State to establish the outer limits "on the basis of" the recommendations of the CLCS, without even stipulating that the coastal State must "take them into account." In essence, the CLCS's role shifted from that of a decision-maker to one that confers legitimacy upon the limits set by the coastal State.[5]

The consequence is that under the original U.S. proposal, the Commission's decision would have been as binding as a ruling by an international judicial body, thus necessitating the sequence of first delineating before proceeding to delimitation, so as to avoid conflicting outcomes. In reality, however, under the UNCLOS, although the CLCS's recommendations are authoritative and effective, the coastal State may deviate from them to some extent. The recommendations simply cannot be equated with a judicial decisions. Moreover, since the UNCLOS treats "delimitation" and "delineation" as two independent processes, and the provisions regarding the CLCS's mandate contain no implication of a required sequence, it might seem logical to assume that the two processes would not interfere with each other regardless of their order. Yet one must ask: if the geographical factors, procedural steps, and legal effects of both "delimitation" and "delineation" remain unchanged, can merely altering the

[1] Bangladesh/Myanmar, supra note 4, at 115, para. 443.

[2] See Question of the Delimitation of the Continental Shelf between Nicaragua and Colombia beyond 200 Nautical Miles from the Nicaraguan Coast (Nicaragua v. Colombia), Preliminary Objections, Judgment, I.C.J. Reports 2016, 17 March 2016, at 132, para.85.

[3] Question of the Delimitation of the Continental Shelf between Nicaragua and Colombia beyond 200 Nautical Miles from the Nicaraguan Coast (Nicaragua v. Colombia), Judgment, Separate opinion of Judge Xue, I.C.J. Reports 2023, 17 March 2016, at 509, para.51.

[4] See Myron H. Nordquist, et al. (2013). *United Nations Convention on the Law of the Sea, 1982: a commentary*, Vol. II. Martinus Nijhoff Publishers, p. 849.

[5] See McDorman, Ted L. (2002). The role of the Commission on the Limits of the Continental Shelf: a technical body in a political world. *The International Journal of Marine and Coastal Law*, *17*(3), 319.

CLCS's role justify claiming that switching the order of the two processes will have no adverse consequences? As noted by Meeting of States Parties to the UNCLOS, coastal states with opposite or adjacent coasts find it difficult to delimit the outer continental shelf before making outer limits final and binding based on the CLCS's recommendations.[1]

4. Intertwined Twins–Where Is the Point of Convergence?

Kunoy argues that "delimitation" and "delineation" are intertwined, and that the latter must precede the former. And failure to follow this fixed sequence would, due to their intertwined nature, affect or interfere with the work of the CLCS.[2] In fact, Kunoy is correct in identifying the intertwined relationship, but he does not delve into the specific elements that connect them.

Only by identifying the precise point where the two processes converge can we analyze—from geographical or legal perspectives—whether a specific sequence should be followed to ensure that one does not prejudice the other. In *Bangladesh/Myanmar Case*, the Tribunal identified this point and drew a series of conclusions: overlapping entitlements are a prerequisite for delimitation, and entitlement should be determined with reference to the outer edge of the continental margin.[3] In the technical-legal process of "delineation," the CLCS is tasked with determining that outer edge of the continental margin. As Judge Ndiaye pointed out in his separate opinion in *Bangladesh/Myanmar Case*, the failure to accurately determine the entitlements to the continental shelf was the greatest shortcoming of that case.[4] "Delimitation" could proceed before "delineation" only if the overlapping entitlements could be precisely determined, not merely deemed "likely" to overlap. Only by employing technical means to present objective geographical data and ascertain the outer edge of the continental margin of both parties can overlapping entitlements be confirmed. Relying solely on a finding of "potential" overlap carries a high risk of error and is not an appropriate method for resolving disputes concerning the limits of national jurisdiction.

Since the delineation is a unilateral act, it must undergo the procedure under Article 76(8) of the UNCLOS to gain acceptance and recognition by the international community—this is the very purpose for which the CLCS was established. However, the UNCLOS's scaling back and compromise regarding the CLCS's powers means that the fundamental technical step of determining the outer edge of the continental margin lacks authoritative resolution. Moreover, coastal States may potentially exaggerate data favorable to them. Materials such as submissions that have not been considered by the CLCS, documents submitted by States in litigation (and even documents agreed by parties, since there may be areas undisputed between them but where a third State has an interest or claim) are generally insufficient to conclusively determine overlapping entitlements to the outer continental shelf.

5. How to Break the Deadlock?

How can the impasse—the failure of "delimitation" and "delineation" to interact constructively—be broken? Recent scholarly views have shed light on this legal issue by refocusing on the fundamental geographical factors that must be considered. Ki Beom Lee, for instance, argues that the sequence should depend on the geographical configuration between the States involved. Between States with opposite coasts where the distance between their coasts is less than 400 nm, delineating the outer limits should take priority, since confirming the natural prolongation of the continental shelf is a "relevant circumstance" that must be considered in delimitation. Where the distance exceeds 400 nm, it is first necessary to determine the relationship between the outer edge of the continental margin and the 200 nm lines of both States to establish whether delimitation is even necessary. Conversely, between States with adjacent coasts, delimitation should come first. This is because the boundary must be established to determine whether both States actually possess a continental shelf

[1] See Issues with respect to article 4 of Annex II to the United Nations Convention on the Law of the Sea Background paper prepared by the Secretariat (SPLOS/64), SPLOS, 1 May 2001, https://docs.un.org/en/SPLOS/64, para.46.

[2] See Kunoy, Bjørn. (2010). The Admissibility of a Plea to an International Adjudicative Forum to Delimit the Outer Continental Shelf Prior to the Adoption of Final Recommendations by the Commission on the Limits of the Continental Shelf. *The International Journal of Marine and Coastal Law*, *25*(2), 262, 270.

[3] Bangladesh/Myanmar, supra note 4, at 105, 114, para. 397, 437.

[4] See Delimitation of the maritime boundary in the Bay of Bengal (Bangladesh/Myanmar), Separate Opinion of Judge Tafsir M. Ndiaye, ITLOS Reports 2012, p. 174, para. 85.

extending beyond 200 nm.[1]

Building on the premise that the relationship between "delimitation" and "delineation" depends on the determination of overlapping entitlements to the outer continental shelf, Ki Beom Lee's perspective offers a geographically-informed approach to understanding the interaction between the two processes. This view accounts for the distinct geographical contexts of States with opposite or adjacent coasts. It notably argues that in the case of adjacent coasts, delimitation should take precedence because the two States may share a single continental shelf. While the outer edge of the continental margin might be determinable, without prior delimitation, it remains unclear whether entitlements overlap or if a State even possesses entitlement to a continental shelf beyond 200 nautical miles. In such scenarios, establishing the outer limits first could allow a party to leverage geographical data to its advantage, potentially prejudicing the subsequent delimitation. For opposite coasts, as mentioned earlier, establishing outer limits first is not mandatory, but it represents the most convincing and expedient approach.

However, this solution is palliative, addressing the symptoms rather than the root cause. While it acknowledges that delimitation should sometimes precede the delineation—aligning with ITLOS's approach of exercising jurisdiction beyond 200 nm only under specific conditions—it fails to resolve the fundamental impediments to the interaction between "delimitation" and "delineation."

The key to breaking this deadlock lies in removing the obstructive clause—Paragraph 5(a)—which impedes the CLCS's consideration of submissions.[2] This clause indirectly grants any of the States concerned in the dispute what amounts to an arbitrary veto power over a submission, requiring no evidence or justification. Some scholars have called for a complete deletion of this clause,[3] while others, along with some states, advocate for its technical modification, such as granting the CLCS the authority to identify the existence of a dispute or to conduct preliminary technical assessments[4].

In the author's view, the most targeted solution is to empower the CLCS to provide preliminary technical assessments on all submissions, irrespective of disputed areas. Firstly, this solution avoids contravening Article 76(10) and Article 9 of Annex II to the UNCLOS, as they prohibit the CLCS to process submissions where sovereignty or delimitation disputes exist. Secondly, this solution aligns with the object and purpose of establishing the CLCS by enhancing its technical role without exceeding its scientific mandate. Last but not least, this solution would effectively address the problem faced by international judicial bodies, which often lack the necessary scientific data to delimit the continental shelf, thereby resolving the dysfunctional interaction between "delimitation" and "delineation."

References

De Herdt, Sandrine W. (2020). The relationship between the delimitation of the continental shelf beyond 200 nm and the delineation of its outer limits. *Ocean Development & International Law*, *51*(3), 11.

Elferink, Alex G. Oude. (2017). Paragraph 5 (a) of Annex I to the Rules of Procedure of the Commission on the Limits of the Continental Shelf: Solution to a Problem or Problem without a Solution?, in Myron H. Nordquist, et al., eds., *Legal Order in the World's Oceans: UN Convention on the Law of the Sea*, Vol. 21, Brill, p. 317.

Kunoy, Bjørn. (2010). The Admissibility of a Plea to an International Adjudicative Forum to Delimit the Outer Continental Shelf Prior to the Adoption of Final Recommendations by the Commission on the Limits of the

[1] Lee, Ki Beom. (2014). Should the Invocation of Paragraph 5 (a) of Annex I to the CLCS Rules of Procedure Result in an Automatic Deferral of the Consideration of a Submission? *Chinese Journal of International Law*, *13*(3), 614-615.

[2] See De Herdt, Sandrine W. (2020). The relationship between the delimitation of the continental shelf beyond 200 nm and the delineation of its outer limits. *Ocean Development & International Law*, *51*(3), 11.

[3] See Elferink, Alex G. Oude. (2017). Paragraph 5 (a) of Annex I to the Rules of Procedure of the Commission on the Limits of the Continental Shelf: Solution to a Problem or Problem without a Solution?, in Myron H. Nordquist, et al., eds., *Legal Order in the World's Oceans: UN Convention on the Law of the Sea*, Vol. 21, Brill, p. 317.

[4] See TRAN, Huu Duy Minh. (2023). The Approach of the Commission on the Limits of the Continental Shelf to Submissions Involving Unresolved Disputes: Should It Be Modified? *Asian Journal of International Law*, *13*(1), 140-143; See Report of the twenty-ninth Meeting of States Parties (SPLOS/29/9), SPLOS, 8 July 2019, https://docs.un.org/en/splos/29/9, para.70.

Continental Shelf. *The International Journal of Marine and Coastal Law*, *25*(2), 262, 270.

Lee, Ki Beom. (2014). Should the Invocation of Paragraph 5 (a) of Annex I to the CLCS Rules of Procedure Result in an Automatic Deferral of the Consideration of a Submission? *Chinese Journal of International Law*, *13*(3), 614-615.

McDorman, Ted L. (2002). The role of the Commission on the Limits of the Continental Shelf: a technical body in a political world. *The International Journal of Marine and Coastal Law*, *17*(3), 319.

Myron H. Nordquist, et al. (2013). *United Nations Convention on the Law of the Sea, 1982: a commentary*, Vol. II. Martinus Nijhoff Publishers, p. 849.

TRAN, Huu Duy Minh. (2023). The Approach of the Commission on the Limits of the Continental Shelf to Submissions Involving Unresolved Disputes: Should It Be Modified? *Asian Journal of International Law*, *13*(1), 140-143.

PIONEER
ACADEMIC PUBLISHING LIMITED

Studies in Law and Justice
ISSN 2958-0382
www.pioneerpublisher.com/slj
Volume 4 Number 5 October 2025

Research on the Soft Law Incentive Mechanism for Coordinated Regional Economic Development

Junjie Zhao[1]

[1] School of Social Sciences, City University of Hong Kong, Hong Kong, China
Correspondence: Junjie Zhao, School of Social Sciences, City University of Hong Kong, Hong Kong, China.

doi:10.56397/SLJ.2025.10.03

Abstract

Against the backdrop of the in-depth advancement of the regional economic coordinated development strategy and the construction of a unified national market, the traditional hard law governance model is difficult to adapt to the complexity and dynamics of cross-regional cooperation due to its rigid constraints. The soft law incentive mechanism, with its characteristics of flexible constraints and consultative governance, has become an important supplement. Based on institutional economics, soft law governance theory and regional rule of law theory, this article systematically analyzes the theoretical basis and functional value of soft law in reducing transaction costs, promoting consultative co-governance and facilitating the synergy of soft and hard laws. Based on China's practice, it is found that soft law faces practical difficulties in regional economic coordination, such as lagging institutional supply, imbalance in the coordination between soft and hard laws, insufficient coverage of norms, and weak implementation efficiency. The research proposes to promote the transformation of the soft law incentive mechanism in regional economic governance from "form innovation" to "efficiency improvement" through paths such as establishing a predictive formulation mechanism, strengthening the transformation of soft and hard laws, and improving the full-cycle management, providing institutional support for deepening regional coordinated development.

Keywords: regional economy, soft law incentives

1. Introduction

Since the Central Economic Conference in 2014, the significance of coordinated regional economic development has become increasingly prominent and has been frequently mentioned. The report of the 20th National Congress of the Communist Party of China has clearly made strategic arrangements, demanding the in-depth implementation of the regional coordinated development strategy, major regional strategies, and the strategy of main functional zones, among others. The 2024 government work report further emphasizes that the key tasks of "promoting urban-rural integration and coordinated regional development" are "deeply implementing the regional coordinated development strategy, major regional strategies, and the strategy of main functional zones, and advancing new urbanization and the comprehensive revitalization of rural areas." That is, unifying the basic market system,

unifying market infrastructure, unifying the standards of government behavior, unifying market supervision and law enforcement, unifying the factor and resource market, and continuously expanding domestic and foreign opening up, this provides fundamental guidance for coordinated regional development.

However, in the process of practice, problems such as interest barriers, policy conflicts and the lack of cooperation mechanisms among various administrative divisions have seriously restricted the depth and effectiveness of regional coordination. The traditional "hard law" governance model characterized by compulsiveness often falls short in dealing with the complexity, flexibility and adaptability of cross-regional affairs. Although regional fiscal and taxation cooperation has a basis for implementation and has formed typical forms such as regional fiscal compensation cooperation, fiscal input cooperation, tax collection and administration cooperation, and tax sharing cooperation, it is still in the stage of policy guidance as a whole, and the level of legal development is obviously insufficient.

In the process of coordinated regional economic development, the limitations of traditional hard law governance have become increasingly prominent, and there is an urgent need for more flexible and adaptable institutional innovations. Soft law, as a rule system formulated by specific entities, featuring written and open characteristics and not relying on state coercive power for implementation, is gradually becoming an important tool to fill the gap of hard law and enhance regional governance efficiency through governance methods such as flexible constraints, incentive guidance, and multi-party consultation. Its mechanism of action is rooted in the dual drive of economic interest incentives and market reputation constraints. It not only encourages business entities to proactively comply with regulations but also allows them to make moderate adjustments to the rules based on their own characteristics, thereby achieving a balance between governance flexibility and order stability.

This article focuses on the incentive value of soft law in regional economic coordination, deeply analyzes the practical predicaments it faces during operation, such as lagging institutional supply, insufficient binding force, and the absence of coordination mechanisms. It also systematically explores the improvement paths of the soft law incentive mechanism from dimensions such as optimizing rule design, strengthening implementation guarantee, and improving supervision and evaluation. With the aim of providing theoretical references and practical guidance for promoting the modernization of regional governance.

2. Theoretical Basis and Functional Value of Soft Law Incentive Mechanism

In the process of coordinated regional economic development, problems such as poor cross-regional resource integration, difficulty in interest coordination, and lagging institutional supply have long existed, and the limitations of the traditional hard law governance model have become increasingly prominent. Against this backdrop, the soft law incentive mechanism has emerged as a key element in promoting the modernization of regional governance, thanks to its unique governance advantages. This mechanism is by no means water without a source; rather, it is deeply rooted in the theoretical fertile soil where multiple disciplines such as institutional economics, governance theory, and regional rule of law interweave. By exerting differentiated functional values, it continuously injects institutional impetus into regional coordinated development.

2.1 Theoretical Basis

From the perspectives of economics, governance theory, regional rule of law, etc., soft law can achieve a balance of multi-level legal structures, specifically including the following three aspects:

Firstly, by drawing on the theories of institutional economics, especially the transaction cost theory and the incentive compatibility theory. From the perspective of transaction cost theory, administrative barriers, policy differences and information asymmetry among regions often make inter-governmental cooperation accompanied by high negotiation, implementation and supervision costs. Soft law can significantly reduce these institutional frictions by formulating standardized cooperation agreements, building information sharing platforms, and establishing collaborative management mechanisms. The theory of incentive compatibility further explains how soft law can reconcile the conflicts of interest among regions. In regional economic cooperation, local governments naturally have the demand to

maximize their own interests. If there is a lack of effective institutional guidance, they are very likely to fall into the "tragedy of the Commons" or the "Prisoner's dilemma". Soft law, through ingenious rule design, organically integrates individual interest demands with regional public interests, enabling local governments to automatically achieve the goal of overall coordinated regional development in the process of pursuing their own growth.

Secondly, by leveraging the theory of soft law governance. The coordinated development of regional economies often involves multiple stakeholders and is confronted with challenges such as a volatile market environment and the frequent emergence of new business forms. The rigid constraints of traditional hard laws are difficult to respond quickly. The "consultative governance" model advocated by the soft law governance theory can precisely break through this predicament. Soft law does not rely on the coercive power of the state. Instead, it uses flexible means such as reputation mechanisms, self-discipline constraints, resource allocation preferences, and policy benefits to guide all parties to actively participate in governance. It not only respects the existing rules and systems of all parties, but also provides guidance for financial business innovation, greatly promoting the integration of regional financial markets. This governance approach based on consultation and aimed at reaching consensus has made soft law an important tool for flexibly responding to complex situations and balancing multiple interests in the coordinated development of regional economies.

Thirdly, draw on the theory of regional rule of law. Regional rule of law theory breaks through the limitation of traditional rule of law that overly relies on hard laws, emphasizing "coordinated governance of both soft and hard laws", providing a macro theoretical support for the role of soft laws in the coordinated development of regional economies. This theory holds that in the process of regional economic integration, complex governance demands can no longer be solely achieved through hard laws. Only by establishing a multi-level and diversified rule system can the stable development of the regional economy be guaranteed. In the practice of coordinated regional economic development, governance scenarios such as maintaining cross-regional market order and the free flow of factors have put forward dynamic and refined requirements for the rule of law. Although hard laws are authoritative and stable, and can establish the basic framework and principles for regional development, they have inherent flaws such as cumbersome formulation procedures and long revision cycles, making it difficult to respond promptly to the governance demands of emerging fields. Soft laws, with their flexible formulation and convenient adjustment, can promptly fill the gaps left by hard laws in the supervision of emerging business forms and regional policy coordination.

From the perspective of regional rule of law theory, the synergy and complementarity between soft law and hard law have demonstrated powerful effectiveness in practice. The flexibility of soft law enables regional cooperation to quickly adapt to market changes and innovative demands, while the enforceability of hard law provides a solid legal guarantee for regional cooperation. The two complement each other and jointly promote the steady and far-reaching development of regional economies on the track of the rule of law, fully demonstrating the practical value of a balanced multi-level legal structure under the guidance of regional rule of law theory.

2.2 Functional Value

The soft law incentive mechanism, with its flexibility, inclusiveness and adaptability, plays an irreplaceable functional value in the coordinated development of regional economies and has become a key institutional innovation for promoting the modernization of regional governance and facilitating the process of regional integration. Its value is specifically reflected in the following four core dimensions.

2.2.1 Fill Institutional Gaps and Drive Innovative Practices

During the rapid transformation of regional economies, new industries are constantly evolving and cross-border integration models are emerging. Traditional hard laws, due to their long formulation cycles and cumbersome revision procedures, are difficult to respond promptly to governance demands. Soft rules, with their advantages of simple formulation processes and flexible adjustments, have become the forefront for filling institutional gaps and incubating innovative paradigms. Take Zhuhai as an example. Relying on the legislative power of the special economic zone, it has introduced a

series of "small-scale" legislations in response to the cross-border cooperation demands of the Guangdong-Hong Kong-Macao Greater Bay Area. Zhuhai was the first to achieve mutual recognition of professional qualifications in fields such as construction, tourism, and healthcare. These soft law practices not only promptly resolved the institutional obstacles to the flow of professional talents in the three regions, but also accumulated practical experience for the subsequent improvement of hard laws, forming an innovative model of "testing soft laws first and establishing hard laws later". Since the implementation of the relevant soft law norms, the number of professional service cooperation projects between Zhuhai and Hong Kong and Macao has increased significantly, providing a solid legal support for the "soft connectivity" of the Greater Bay Area and fully demonstrating the role of soft law as a "test field" in institutional innovation.

2.2.2 Break down Cooperation Barriers and Build a Collaborative Network

In traditional regional cooperation, the rigid constraints of hard laws often make local governments cautious about cross-regional cooperation due to concerns over responsibility risks and imbalance of interests. The non-coercive nature of soft law has broken this deadlock. Through flexible rules of consultation and dialogue as well as interest coordination, it has lowered the threshold for cooperation and stimulated the enthusiasm of local governments to participate in collaborative governance. Take the inter-provincial cooperation mechanism of the Yangtze River Economic Belt as an example. The provinces and cities along the route have signed soft law nature cooperation agreements such as ecological compensation and industrial transfer, clarifying the rights, responsibilities and interests of all parties. On the basis of respecting local autonomy, they have built a collaborative network of resource sharing and risk sharing. According to statistics, after the implementation of the relevant soft law agreements, the ecological joint prevention and control of cross-regional industrial cooperation projects in the Yangtze River Economic Belt has achieved remarkable results. This consultative governance model based on soft law has effectively resolved potential conflicts among regions and promoted the formation of a new pattern of cooperation that has shifted from "passive response" to "active coordination".

2.2.3 Stabilize Market Expectations and Stimulate Economic Vitality

In terms of optimizing the regional business environment and stimulating market vitality, the Soft Law, by formulating industry standards, best practice guidelines and other normative documents, provides clear behavioral guidance for market entities and enhances their confidence in cross-regional investment and operation. Zhuhai has passed the "Regulations on the Administration of the Registration of Market Entities in the Zhuhai Special Economic Zone", clearly stating that it will cooperate with service institutions in Hong Kong and Macao to achieve "cross-border processing" for the registration of market entities, significantly reducing the time for Hong Kong and Macao enterprises to settle in Zhuhai. The district (comprehensive) legal affairs zone established in Shifang City, Deyang, has adopted soft law measures such as releasing enterprise compliance guidelines and forming a legal service alliance. Providing full life-cycle legal support for enterprises effectively reduces their operational risks. Data shows that after the establishment of the Legal Affairs Zone, the number of contract disputes among enterprises in Shifang City has dropped significantly, and the satisfaction with the business environment has greatly improved. It has attracted a large number of enterprises to invest and settle down, effectively playing the role of soft law in stimulating market vitality.

2.2.4 Reconcile Differences in Rules and Promote Regional Integration

In regions with significant differences in rule systems, such as the Guangdong-Hong Kong-Macao Greater Bay Area, soft law has become an important bridge to eliminate institutional barriers and promote rule integration. Through soft law interaction mechanisms such as industry association exchanges and expert think tank discussions, market entities and regulatory authorities from different jurisdictions have been able to communicate deeply and gradually reach a consensus on rules. In the field of intellectual property protection, the Greater Bay Area has unified patent examination standards and trademark registration processes through soft law measures such as formulating joint standards and establishing mediation and

collaboration mechanisms, significantly shortening the cross-regional intellectual property dispute resolution cycle. In terms of data flow, the Soft Law has promoted the establishment of a mutual recognition mechanism for data security assessment in the Greater Bay Area, facilitating the orderly circulation of data elements. This continuous dialogue and integration of rules not only promotes the free flow of factors within the region but also lays an institutional foundation for building a unified and open regional market, accelerating the process of regional integration.

3. The Practical Forms and Predicaments of Soft Law Incentives in Regional Economic Synergy in China

Under the new normal of the economy characterized by "medium-high growth, deep structural adjustment and innovation-driven development", China's regional economic coordinated development strategy is undergoing a profound transformation from local breakthroughs to all-round linkage. The soft law incentive mechanism, with its flexible and consultative characteristics, plays a significant role in regional cooperation. However, many deep-seated contradictions have also been exposed in practical exploration. These predicaments not only restrict the effectiveness of the regional coordinated development strategy but also put forward urgent requirements for institutional optimization. Specifically, they can be analyzed from two aspects: structural defects at the institutional formulation level and functional obstacles in the implementation process.

4. Practical Predicament: The Dual Challenges of Institutional Supply and Implementation Effectiveness

4.1 The Deep-Seated Contradictions in the Formulation of Regional Coordinated Development

(1) Insufficient forward-looking and lagging dynamic adaptation

Against the backdrop of the accelerated iteration of regional development strategies, the soft law normative system shows a significant lag. At present, the vast majority of regional soft law agreements are still concentrated in traditional infrastructure, energy cooperation and other fields, while the supply of rules for emerging industries such as digital economy, green finance and cross-border e-commerce is seriously insufficient.

(2) Structural imbalance in the coordination mechanism between soft and hard laws

In the governance of regional economy in our country, soft laws and hard laws have not yet formed a virtuous interactive ecosystem. On the one hand, the absence of a review mechanism for normative documents has led to frequent occurrences where soft laws overstep their bounds or conflict with hard laws. For instance, a certain western province has, in its regional industrial support policies, added local protection provisions in the form of soft laws that are contrary to the Anti-Monopoly Law, thus undermining the fair market competition environment. On the other hand, in emerging fields where hard law is absent, soft law has failed to fully play its role as a "testing ground". In cutting-edge fields such as artificial intelligence ethics and carbon footprint accounting, the practice of soft laws mostly remains at the level of principle-based advocacy and fails to promote the improvement of hard laws through effective transformation, thus missing the golden window period for institutional innovation.

(3) Deficiencies in the coverage and formal rationality of the normative system

There are obvious problems of "incomplete coverage of fields" and "process discontinuity" in the soft law norms of regional economy. From the perspective of the field dimension, the norms for industrial chain collaboration mostly focus on the production and manufacturing links, with insufficient coverage of upstream and downstream links such as R&D and design, supply chain management, and after-sales operation and maintenance, resulting in breakpoints in regional industrial collaboration. From the perspective of process, key links such as project approval and operation supervision lack systematic design. For instance, in the construction norms of industrial parks, detailed rules for ecological protection during the construction process and operation performance evaluation are generally absent. At the same time, the norms themselves have the drawbacks of ambiguous expression and loose logic. Although the logistics collaboration agreement of a certain region proposed "promoting information sharing", it did not clearly define the data interface standards, sharing scope and security responsibilities, which made it difficult for the information systems of enterprises to be interconnected and weakened the operability of

the rules.

4.2 Practical Obstacles in the Implementation of Regional Coordination

4.2.1 Non-Enforceability Leads to a Performance Crisis

The non-coercive nature of soft law poses a severe test to it in the game of interests. In the field of regional fiscal and taxation cooperation, the number of disputes over the performance of soft law agreements has grown rapidly on average over the past three years, with the vast majority of them stemming from the absence of liability clauses for breach of contract. A certain energy supply guarantee agreement in the western region failed to clearly define a price fluctuation compensation mechanism. When the market price of coal rose, the supplier unilaterally defaulted, causing three listed companies downstream to suspend production and resulting in significant direct economic losses. This highlights the fatal impact of low default costs on the stability of cooperation.

4.2.2 Fragmented Practices Dissolve the Synergy Effect

The current soft law mechanism shows a significant feature of "isolation". During the process of transportation integration in the Beijing-Tianjin-Hebei region, although the railway, highway and aviation departments have signed 28 cooperation agreements, due to the lack of a unified dispatching platform, the efficiency of multimodal transport connection is far lower than that of developed countries, and at the same time, the logistics cost has increased compared with expectations. This governance model of "each managing its own section" makes it difficult for soft laws to form a policy synergy, and the efficiency of regional resource integration is greatly reduced.

4.2.3 The Absence of a Benefit Compensation Mechanism Weakens the Impetus for Cooperation

The contradiction of interest redistribution in regional collaboration has become increasingly prominent. In the industrial transfer of the Yangtze River Economic Belt, the average industrial tax revenue of the transferring-out areas was lost due to the relocation of enterprises. However, the receiving areas generally failed to establish a tax-sharing and employment compensation mechanism, resulting in an increase in cases where the transferring-out areas set up hidden barriers to hinder the relocation of enterprises compared with the previous year. The contradiction in the field of ecological compensation is more acute. The upstream provinces invest over 100 billion yuan annually in environmental protection, but the transfer payments received by the downstream are relatively low, which directly affects the enthusiasm for ecological protection.

4.2.4 Loopholes in Full-Cycle Management Hinder the Iteration of the System

The problem of the lack of closed-loop management in the formulation and implementation of soft laws is serious. During the formulation stage, the participation rate of the public and enterprises was insufficient. As a certain science and technology innovation cooperation agreement in the Greater Bay Area failed to incorporate the opinions of small and medium-sized enterprises, most of its provisions were difficult to be implemented in practice. The implementation stage is even more lacking in a dynamic monitoring system. A certain inter-provincial innovation fund, due to the lack of setting assessment indicators for the transformation of scientific and technological achievements, has had an insufficient utilization rate of fiscal funds for three consecutive years, yet it has been unable to adjust policies in a targeted manner, resulting in an inefficient cycle of "formulation—failure—re-formulation".

5. The Path to Improving the Soft Law Incentive Mechanism in the Coordinated Development of Regional Economies

5.1 Solutions to Challenges at the Institutional Supply Level

5.1.1 Establish a Regional Development Frontier Trend Monitoring Center

Integrate resources from universities, think tanks, and industry associations, continuously track technological evolution and business model innovation in emerging fields such as the digital economy, green industries, and artificial intelligence, and regularly release the "White Paper on Rules and Demands for Regional Coordinated Development" to provide forward-looking guidance for the formulation of soft laws.

Promote "rolling" soft law planning: Change the fixed medium and long-term planning model and implement a rolling revision mechanism of "three years of core provisions + two years of

dynamic adjustment". The implementation effect and applicability of the soft law agreement are evaluated annually, and the revision process is initiated in a timely manner to ensure that the system is in line with economic dynamics.

Prioritize the implementation of "soft law pilot projects" in emerging fields: For areas not yet covered by national hard laws such as cross-border data and artificial intelligence ethics, encourage pilot regions (such as free trade zones and new areas) to take the lead in introducing detailed soft law guidelines, accumulate experience, and provide "local samples" for national legislation.

5.1.2 Addressing the Structural Imbalance in the Synergy Mechanism of Soft and Hard Laws

Establish dual channels for soft law compliance review and hard law conversion: Set up a soft law filing and review mechanism. All soft law agreements signed between regions must be submitted to the superior legislative body (such as the Standing Committee of the People's Congress) or relevant central government functional departments (such as the National Development and Reform Commission and the Ministry of Justice) for filing and compliance review to ensure that they do not conflict with higher-level laws (such as the Anti-Monopoly Law), and to prevent "overreach of soft law" from the source.

Establish a "soft law—hard law" transformation and certification mechanism: For soft regulations that have been tested in practice, proven effective and have a high degree of consensus, establish mature recognition and transformation procedures. The core content is absorbed and refined by the superior legislative body, and then transformed into local regulations, departmental rules or even national laws through legal procedures, thus breaking through the "last mile" of institutional innovation.

5.1.3 Promote the Rule Design of "Full Industrial Chain" and "Full Life Cycle"

Expand the breadth of rule coverage: When formulating the soft law for industrial collaboration, it is required to cover the entire chain of "R&D—design—production—supply chain—marketing—after-sales service", especially to make up for the shortcomings such as R&D cooperation, intellectual property sharing, supply chain resilience, and unified after-sales service standards.

Deepen the depth of rules and processes: For project approval, operation supervision and other links, issue supporting detailed implementation rules and standard contract texts. For instance, in the construction norms of industrial parks, it is mandatory to incorporate environmental protection provisions, energy consumption standards, performance evaluation methods and exit mechanisms.

Enhance the accuracy of soft language texts: Promote the use of standardized, indexed, and quantified protocol languages. Avoid using vague words such as "strengthen", "promote" and "improve", and clearly define the subject, standards, procedures, time limits and responsibilities. For instance, "promoting information sharing" should be specified as "sharing [a certain type of data] within [a specific scope] based on [specific international/national standards] API interfaces, and the [designated party] assumes the responsibility for data security."

5.2 Solutions to Challenges at the Implementation Efficiency Level

5.2.1 Innovate the Soft Law Performance Guarantee Mechanism

Introduce "flexible constraints" tools: Generally, include terms such as breach of contract guarantee, reputation notification, and loss of priority cooperation rights in the agreement. Establish cross-regional credit files for enterprises and local governments, incorporate major breach of contract behaviors into credit records, and link them to fiscal transfer payments, project approval, official performance evaluations, etc. Promote "standardized + personalized" agreement templates: Led by the central government, standard agreement templates should be formulated for high-frequency cooperation areas such as energy supply security and ecological compensation. These templates must include price linkage mechanisms, risk-sharing clauses, and clear methods for determining liability for breach of contract and calculating compensation.

5.2.2 Build an Implementation Framework for "Holistic Governance"

Establish a substantive and permanent regional coordination body: Go beyond the "joint conference" model and set up a permanent office with certain administrative coordination and resource allocation powers (such as the "Office for Coordinated Development of the

Beijing-Tianjin-Hebei Region"), responsible for coordinating the implementation of various soft laws and coordinating the resolution of cross-departmental and cross-regional conflicts.

Build a unified digital collaborative platform: Create a regional "digital hub" covering functions such as multimodal transport, data sharing, and environmental monitoring. It is mandatory for all signatories to integrate relevant data and processes into the platform, breaking down "islands" through technical means to achieve the automation and visualization of business collaboration.

5.2.3 Design a Scientific and Reasonable Interest Balance and Compensation System

Establish a "horizontal fiscal transfer payment" system: In industrial transfer, calculate the tax-sharing ratio between the transferring-out area and the receiving area based on a clear formula (for example, share a certain proportion of the taxes generated by the relocated enterprises in the first five years). The provincial or central government is responsible for the transfer of funds to ensure their implementation.

Improve the market-oriented and diversified ecological compensation mechanism: Comprehensively utilize various tools such as fiscal transfer payments, off-site development rights transactions, water resource transactions, and green financial products. Establish an accounting system based on the value of ecosystem services to match the compensation standards with the protection costs and benefits, ensuring that the upstream regions have a sense of "gain".

5.2.4 Improve the Closed-Loop System of "Formulation—Implementation—Evaluation—Revision" for Soft Laws

Strengthen multi-party participation in the formulation process: It is mandatory to stipulate that the formulation of soft laws must go through procedures such as public announcement, hearing, and expert argumentation, especially to involve small and medium-sized enterprises, industry associations, and public representatives. The participation situation should be one of the preconditions for the effectiveness of the agreement.

Establish a quantitative performance monitoring and dynamic evaluation mechanism: Set key performance indicators (KPIs) for each important soft law agreement, such as the efficiency of fund utilization, project implementation rate, and enterprise satisfaction. The assessment report is regularly evaluated by a third-party institution and publicly released.

Establish a rigid feedback and iterative mechanism: Based on the assessment results, for the agreement terms with poor implementation effects, it is stipulated that the revision process must be initiated within the prescribed time limit (such as within half a year), forming a virtuous cycle of "assessment—feedback—revision—re-implementation", and eliminating the phenomenon of "only formulating but not managing".

6. Conclusions

The coordinated development of regional economies is an important engine for high-quality economic development in China in the new era, and the improvement of the soft law incentive mechanism is the key path to solving the problems in regional governance. From a theoretical perspective, soft law has constructed a three-dimensional analysis framework of "cost—benefit—rule of law" through the integration of multi-disciplinary theories, providing a governance logic distinct from that of hard law for regional collaboration. From a practical perspective, although the current application of soft law still faces the dual challenges of institutional supply and implementation efficiency, in emerging fields such as data element circulation and ecological compensation, soft law has demonstrated its unique value in filling institutional gaps.

Looking to the future, the governance of regional economic soft laws needs to achieve three breakthroughs: First, in terms of rule supply, it should shift from "passive response" to "forward-looking guidance", and enhance the adaptability of soft laws to cutting-edge fields such as the digital economy through dynamic monitoring and rolling revision mechanisms. Second, in terms of governance logic, it shifts from "single soft law" to "soft and hard synergy", and builds a clearly hierarchical regional legal system through compliance review and transformation certification mechanisms. Third, in terms of implementation efficiency, it shifts from "fragmented governance" to "holistic collaboration". Through digital platforms and benefit compensation mechanisms, soft regulations are transformed into tangible regional cooperation

momentum.

With the in-depth advancement of the "Five unifications and One Opening-up" requirements, the soft law incentive mechanism will play a more crucial role in breaking down administrative barriers, promoting the flow of factors, and fostering a unified market. This requires not only a continuous deepening of the understanding of the laws of soft law governance at the theoretical level, but also the transformation of the advantages of soft law into governance effectiveness through institutional innovation in practice. Ultimately, it aims to form a regional economic collaborative soft law governance model with Chinese characteristics, providing an institutional innovation sample for global regional governance.

References

Hui Yan. (2025). Research on the Correlation Between Fiscal and Tax Policies and Regional Economic Coordination. *Marketing of Time-honored Brands*, *05*, pp. 85-87.

Li Nan. (2020). Research on Soft Law Governance for Regional Economic Coordination (Doctoral Dissertation, Southwest University of Political Science and Law). Doctoral. http://gfagz304790044ec44369svb99kv56vpnf6ouv.fzfy.oca.swupl.edu.cn/10.27422/d.cnki.gxzfu.2020.001758.

Ma Guilan. (2012). On the Policy Coordination Mechanism of Regional Economic Development. *Journal of Henan Normal University (Philosophy and Social Sciences Edition)*, *39*(02), 112-115. http://gfagz304790044ec44369svb99kv56vpnf6ouv.fzfy.oca.swupl.edu.cn/10.16366/j.cnki.1000-2359.2012.02.055.

Sun Jiwen & Jiang Zhi. (2025). Systems, Mechanisms, and Policies for High-Quality Coordinated Regional Economic Development in China During the 15th Five-Year Plan Period. *Economic Theory and Economic Management*, *45*(07), 1-14.

Zhang Yanhua. (2014). Research on Regional Economic Coordinated Development from an Industrial Perspective. Doctoral Dissertation, Wuhan University. http://gfagzcadd5f6184ce4461svb99kv56vpnf6ouv.fzfy.oca.swupl.edu.cn/kcms2/article/abstract?v=Axm2Bq0K5ftmBHJnn9sPkE6YzbsjLRTGshfklxtrfVBkP_ysJ9_GOh68BOUVqYoFQNScrZAVLi4dkd2bdn_WC5rcTTUQ21xNgoL4asIsjLHDaK9XpJ5qByND8x2qnpZp4HrfTAloNd-F-C3_CCm_HnDnSjjOrDB6aVvZ-EVF48SYWicrRL-KQFhS3oe2qYAS&uniplatform=NZKPT&language=CHS

PIONEER
ACADEMIC PUBLISHING LIMITED

Studies in Law and Justice
ISSN 2958-0382
www.pioneerpublisher.com/slj
Volume 4 Number 5 October 2025

Does International Humanitarian Law Adequately Address Gender Issues?

Ruixuan Wu[1]

[1] The University of Sydney, Australia
Correspondence: Ruixuan Wu, The University of Sydney, Australia.

doi:10.56397/SLJ.2025.10.04

Abstract

With the increasing complexity of armed conflict situations around the world, how to solve the gender issue in conflict situations has gradually become a topic of concern for the international community. International humanitarian law (IHL), as a legal framework designed to provide humanitarian protection for people in armed conflict, is of great significance in regulating wrongful acts in armed conflict and reducing the adverse effects caused by armed conflict, which of course also includes dealing with gender issues that commonly occur in conflict situations. This essay aims to study the significance and problems of gender provisions in IHL, and how different cultural, legal and political factors affect the implementation of gender provisions in IHL, especially the difficulties in the accountability and punishment of sexual violence, from the perspective of the development of gender regulations under the framework of IHL. This paper mainly adopts a qualitative analysis method and supplemented with appropriate quantitative data. Through case studies of specific conflict countries, comparative analysis and analysis of key legal texts under the framework of IHL, this essay expounds the current achievements and shortcomings of IHL in addressing gender issues in armed conflict. The significance of this study is that it hopes to further understand the IHL framework from a gender perspective and put forward relevant suggestions to promote the appropriateness and effectiveness of IHL as an important humanitarian gender protection legal system.

Keywords: IHL gender provisions, gender issues, humanitarian gender protection

1. The Historical Evolution of Gender Regulation Under the IHL Framework

In armed conflicts, gender often has a differentiated impact on the gender. Among other things, women and girls not only face the challenges of inequality in long-term social, economic and political systems, but also the structural gender differences are more pronounced in conflict situations. In a survey conducted by the United Nations High Commissioner for Refugees (UNHCR) with Syrian refugee women, most of whom had fled to other countries, they were often forced to live in poor and overcrowded housing without basic necessities, and did not have enough money to support their families.[1] Syrian refugee women who face the loss of adult male company are even more vulnerable. In communities with inadequate infrastructure and overcrowding,

[1] United Nations High Commissioner for Refugees. (2014). *Woman Alone: The Fight for Survival by Syria's Refugee Women*. Report, 15.

women are not only marginalised in the process of obtaining assistance, but are also often subject to abuse.[1] This reflects the many challenges faced by women in conflict situations in terms of access to economic resources and basic services. Women are also underrepresented in decision-making bodies. A report on women's political participation published by UN Women in 2023 shows that the average proportion of women in parliament worldwide is only 26.5% in both the upper and lower houses. At the same time, in terms of holding the highest positions in the country, the proportion of countries with women as head of state and head of government is only 11.3% and 9.8% respectively.[2] From these aspects, it is not difficult to see that military operations have a more serious impact on women.[3]

With the changing forms of international armed conflict and people's concerns about gender-based issues in war, IHL has developed some provisions dedicated to addressing gender issues based on the protection of civilians and prisoners of war in general, which is evident in many articles of the Geneva Conventions. The Geneva Conventions clearly stipulate that when members of the armed forces of a belligerent country, other legal combatants, and religious and medical personnel with the status of prisoners of war, etc. are captured, they have a series of rights and certain obligations to the belligerent country that has detained them, and Article 12-16 of the Third Geneva Convention on the general protection of prisoners of war is the core content of the Convention.[4] Among them, article 16, when discussing the issue of equal treatment of prisoners of war by the detaining State, emphasizes that the detaining State should take into account the provisions of the Geneva Conventions on rank and sex.[5] This means that the article emphasizes the principle of equality to ensure that prisoners of war will not be unfairly treated because of race, religion, political beliefs and other factors, adding consideration to the protection of gender issues. At the same time, other provisions of the Third Geneva Convention embody gender-specific protection in detail, taking into account that women prisoners may have special needs and face special risks while in detention, which may be due to the special physical and biological needs of women. It may also be due to concerns about differences in social, economic, cultural, and political structures between countries and regions.[6] For example, the provisions of article 14, paragraph 2, of the Third Convention take full account of the special needs of women (such as pregnancy and childbirth), prohibit any unequal treatment of women in the status of prisoners of war, and require that women enjoy equal treatment with their sex and equal treatment with men at the preferential level. This has positive implications for the special protection given the increasing number of women in active combat roles in the armed forces of States following the adoption of the Third Convention.[7] In terms of the provisions on the housing of prisoners of war, in the comprehensive revision of the treatment of prisoners of war in the Third Convention after the Second World War, the current article 25, paragraph 4, was introduced, and the detaining State was responsible for implementing the requirement that male and female quarters must be separated, and male prisoners of war could not enter the female quarters regardless of whether they had obtained the consent of female prisoners of war.[8] The purpose is to ensure the special needs and safety of female prisoners of war. With regard to health facilities, article 29 of the third Convention, while requiring the detaining State to take the necessary health measures with regard to prisoners of war to prevent diseases in prison camps and to respect the human dignity of prisoners of war,

[1] Ibid 37.

[2] Inter-Parliamentary Union and UN Women. (2023). *Women in Politics: 2023*. Report, 1.

[3] International Committee of the Red Cross, 'Gender and IHL', (Web Page, 2024) <https://www.icrc.org/zh/law-and-policy/gender-and-ihl>.

[4] Robert Kolb and Richard Hyde. (2008). *An Introduction to the International Law of Armed Conflicts*. Bloomsbury Publishing, 209.

[5] *Geneva Convention Relative to the Treatment of Prisoners of War*, opened for signature 12 August 1949, 75 UNTS 135 (entered into force 21 October 1950) art 16.

[6] International Committee of the Red Cross, 'Commentary of 2020: Article 16 - Equality of Treatment of Prisoners', (Web Page, 2020). <https://ihl-databases.icrc.org/zh/ihl-treaties/gciii-1949/article-16/commentary/2020#12>.

[7] International Committee of the Red Cross, 'Commentary of 2020: Article 14 - Respect for the Persons and Honour of Prisoners', (Web Page, 2020). <https://ihl-databases.icrc.org/en/ihl-treaties/gciii-1949/article-14/commentary/2020?activeTab=>.

[8] International Committee of the Red Cross, 'Commentary of 1960: Article 25 - Prisoners of War and Quarters', (Web Page, 1960). <https://ihl-databases.icrc.org/en/ihl-treaties/gciii-1949/article-25/commentary/1960?activeTab=>.

specifically provides for the provision of separate facilitation measures for women prisoners of war in prison camps.[1] This not only helps to prevent women prisoners of war from being subjected to gender-based violence, but also enables them to enjoy certain privacy rights and specific protections during detention. With regard to the provisions on the Labour of prisoners of war, article 49 (1) of the Third Convention, when arranging Labour on the basis of individual conditions and determining the suitability of prisoners of war for work, not only takes into account the differences in their age, rank and physical ability to adapt their Labour to their individual circumstances, but also takes into account gender.[2] Although the article did not explicitly specify the difference between the contents of the work of female prisoners of war, it required that women prisoners of war should be arranged to avoid imposing inappropriate contents of work on them in order to protect their physical and mental health. In the enforcement of penalties, a number of provisions of the Third Convention involve special consideration of gender protection. For example, article 88 (2), of the Third Convention stipulates that the punishment of female prisoners of war shall not exceed the punishment of female members of the armed forces of the detaining State for the same offence, while paragraph 3 of the same article further optimizes the minimum guarantees for female prisoners of war by prohibiting the punishment of female prisoners of war to be more severe than that imposed on male members of the armed forces of the detaining State for similar offences.[3] This fully interprets the requirement of Article 14 (2), of the Convention that women shall enjoy the same preferential treatment as men in all circumstances, and expresses the prohibition of discrimination on the basis of sex under IHL. Article 94 (4) requires that women prisoners of war who are subject to disciplinary punishment must simultaneously meet two strict protection measures, they must be kept in separate places from male prisoners of war and must be directly supervised by female guardians.[4] This provision is similar to article 108 (2), of the Convention, which applies to the need for separate accommodation and guardianship by specific personnel for women subject to disciplinary action, and obliges the detaining State to comply with obligations stemming from respect for the person and honour of women, with an explicit prohibition of sexual violence on the basis of gender considerations.[5] To some extent, the special needs of female prisoners of war who were sentenced to imprisonment were met.

In addition to the special protection of women in armed conflict provided for in the Third Geneva Convention, the requirement to respect the special needs of women is also reflected in the Fourth Geneva Convention and Additional Protocol I to the Geneva Conventions, in particular with regard to the special care given to pregnant women and mothers of young children. For example, with regard to the choice of targets for attack by parties to a conflict, article 18 of the fourth Convention stipulates that civilian hospitals organized to care for pregnant women, the wounded and the sick shall under no circumstances be the target of attack and shall continue to be protected and respected by parties to a conflict.[6] At the same time, article 21 of the Convention, in keeping with the principle of respect for women enshrined in article 18, requires that convoys or hospital trains on land, as well as ships at sea, should not be targeted by parties to a conflict while transporting pregnant women, civilians and the vulnerable.[7] With regard to the preferential treatment of non-repatriated persons, article 38 of the fourth Convention provides, in particular, criteria for children under 15 years of age, pregnant women and mothers of children under 7 years of age to enjoy the same categories of preferential

[1] International Committee of the Red Cross, 'Commentary of 2020: Article 29 - Hygiene', (Web Page, 2020). <https://ihl-databases.icrc.org/en/ihl-treaties/gciii-1949/article-29/commentary/2020?activeTab=>.

[2] International Committee of the Red Cross, 'Commentary of 2020: Article 49 - General Observations on Labour of Prisoners of War', (Web Page, 2020). <https://ihl-databases.icrc.org/en/ihl-treaties/gciii-1949/article-49/commentary/2020?activeTab=>.

[3] *Geneva Convention Relative to the Treatment of Prisoners of War*, opened for signature 12 August 1949, 75 UNTS 135 (entered into force 21 October 1950) art 88.

[4] Ibid art 97.

[5] International Committee of the Red Cross, 'Commentary of 2020: Article 108 - Execution of Judicial Penalties: Premises and Essential Safeguards', (Web Page, 2020). <https://ihl-databases.icrc.org/en/ihl-treaties/gciii-1949/article-108/commentary/2020>.

[6] *Geneva Convention Relative to the Protection of Civilian Persons in Time of War*, opened for signature 12 August 1949, 75 UNTS 287 (entered into force 21 October 1950) art 18.

[7] Ibid art 21.

treatment as the local population.[1] This is also a reflection of the special care given to women during wartime. Additional Protocol I to the Geneva Conventions also focus on the relief and protection of women in need of special care. For example, article 70 (1) of the Protocol requires that priority be given to persons entitled to special treatment and protection, such as pregnant women, women who have given birth and nursing mothers, in the distribution of relief supplies.[2] At the same time, article 76 of the Protocol expressly prohibits all forms of rape, forced prostitution and indecent assault against women, gives special treatment (such as early release) to pregnant women and mothers raising minor babies who are detained during armed conflict, and the parties to the conflict shall use their utmost efforts to avoid the imposition of the death penalty on such women for crimes committed as a result of armed conflict.[3] To some extent, this is a development and extension of the relevant provisions of the Fourth Convention, which not only protects women in need of special care in the course of armed conflict, but also takes into account the respect and basic guarantees for women in war in general.

Customary International Humanitarian Law (CIHL) is a collection of legally binding rules that are generally observed by States, even if they are not formally enshrined in international treaties. It is also an important part of IHL because of the high degree of consistency with which it is applied by States in armed conflict. For example, Article 93 of the CIHL on the prohibition of rape and other forms of sexual violence not only summarizes the provisions of the Fourth Geneva Convention and the Additional Protocol I to the Geneva Conventions on the prohibition of sexual violence such as rape and forced prostitution, but also indicates that many military manuals clearly define this act as a war crime. It also provides domestic precedents of corresponding countries to support the prohibition of sexual violence in armed conflict, such as the "Takashi Sakai case" tried by the War Crimes Military Tribunal of the Ministry of National Defense of China in 1946, which confirmed that the crime of rape was included in war crimes.[4] At the same time, the Rules refine the definition of rape from the specific cases of the International Criminal Court, not only taking into account the determination of the accurate scope of rape under the framework of international law in the case law of Kunarac case in 2001, but also indicating that the prohibition of sexual violence under IHL is non-discriminatory in practice. That is, both men and women, adults and children are equally protected against sexual violence. For example, except for the act of forced pregnancy, the Statute of the International Criminal Court defines the object of sexual violence as "any person", that is, not limited to women.[5] This shows that CIHL not only has great significance in defining the scope of application of gender protection, but also fills the gap on gender provisions in international treaties to a certain extent.

2. The Need for Gender-Specific Protection in the IHL Framework: The Implications of Differentiated Gender Regulation

Up to now, the pattern of international armed conflicts is still not optimistic. In a speech at Columbia Law School in 2023, International Committee of the Red Cross (ICRC） President Mirjana Spoljaric mentioned that the ongoing global armed violence in more than 100 countries, especially the international armed conflict between Russia and Ukraine, has greatly affected the global political landscape. This has resulted in the destruction of national infrastructure and the suffering of the people for a long time, and these conflicts are mainly manifested in the oppression and persecution of race, religion, class and gender, among which, wherever conflicts exist, the equal rights of women and girls will be reduced and destroyed to varying degrees.[6] A 2021 United Nations report states that in countries experiencing armed conflict, the proportion of households

[1] International Committee of the Red Cross, 'Commentary of 1958: Article 38 - Non-repatriated persons I. General observations', (Web Page, 1958). <https://ihl-databases.icrc.org/en/ihl-treaties/gciv-1949/article-38/commentary/1958>.

[2] *Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I)*, signed 8 June 1977, 1125 UNTS 3 (entered into force 7 December 1978) art 70.

[3] Ibid art 76.

[4] Jean-Marie Henckaerts and Louise Doswald-Beck. (2005). *Customary International Humanitarian Law Volume I: Rules.* Cambridge University Press, 324–25.

[5] Ibid 325–26.

[6] Mirjana Spoljaric. (2023, March 3). *Gender Equality and War: "No Humanity, Dignity, Peace Until International Humanitarian Law Is Upheld for All Genders".* Speech, Columbia University Law School.

headed by women usually reaches 33%, only 44% of women are likely to be engaged in paid work, and the proportion of women in parliament is only 18.9%, especially in countries affected by conflict for a long time, these proportions will be even lower.[1] This means that women face more challenges than men in armed conflict on various levels, limiting their right to fair treatment. Since conflict situations have different manifestations in terms of their impact on gender-related issues, the question of how to coordinate and address gender protection issues within the framework of IHL is an urgent topic.

In situations of armed conflict, resource scarcity often has a severe impact on the civilian population, particularly the relatively vulnerable. In such Settings, women have a greater need for adequate food to safeguard their health and well-being, but they often have a lower status in the social structure as a vulnerable group, which means they are more vulnerable to conflict and have fewer resources.[2] For example, there are traditional practices in some countries that give preference to men in terms of food distribution, while limiting the amount of food consumed by women.[3] In addition, men in armed conflict are often displaced from their families and communities by being mobilized to fight, suffer casualties or are displaced, which prevents them from continuing to do the farming, fishing and paid Labour for which they are responsible to compensate their families.[4] This means that women in times of armed conflict often assume the role of the head of the household, not only to feed themselves, but also to take care of the household and to buy food, which puts more pressure on already financially difficult families and can create a cycle of poverty. Lack of resources can have other serious consequences in very dangerous conflict countries. For example, in the Democratic Republic of the Congo, where armed conflict intensifies and resources are scarce, women and girls may be exposed to targeted violence, abuse and exploitation as they seek income opportunities and food resources due to very limited civil protection and security measures.[5] On the one hand, this indicates that there are potential gender inequality and discriminatory norms in some countries before the adverse situation for women occurs, and on the other hand, it also reflects the insufficient gender protection under the framework of IHL.

In the context of humanitarian assistance to civilians in countries in armed conflict, men have been the primary targets of humanitarian resource planning and delivery. Furthermore, aid agencies do not meet accurate and systematic standards in collecting detailed data on the gender and age characteristics of recipients of assistance from the local civilian population and refugee camps in conflict countries (e.g. households headed by women are prone to omission and neglect). The lack of such data can lead to a lack of comprehensive understanding of the real needs of different gender groups in the planning and specific distribution of humanitarian assistance, which challenges women in an already vulnerable position to obtain adequate and appropriate humanitarian assistance.[6] The conflict in Syria, for example, has greatly affected women's ability to enjoy their most basic rights. As a result of the massive displacement caused by war and the increase in the number of female heads of household, 74 per cent of the millions of civilians in need of assistance are women and girls, and female-headed households are more likely to be unable to meet basic needs than those with male heads of household, with women who have been displaced for long periods of time being disproportionately affected.[7]

While men are at higher risk of war injury or death due to their higher participation in armed conflict, women are generally more vulnerable

[1] International Committee of the Red Cross. (2022, June). *Gendered Impacts of Armed Conflict and Implications for the Application of International Humanitarian Law*. Report, 11.

[2] *Inter-Agency Workshop on Integration of Gender into Needs Assessment and Planning of Humanitarian Assistance.* (1999, July). Summary Report, 1.

[3] Charlotte Lindsey, 'Women Facing War', (Web Document, October 2001), 78 <https://www.icrc.org/sites/default/files/external/doc/en/assets/files/other/icrc_002_0798_women_facing_war.pdf >.

[4] Ibid.

[5] World Food Programme, 'Women, Girls and the Protection Crisis in Eastern DRC', (Web Page, 7 March 2024). <https://reliefweb.int/report/democratic-republic-congo/women-girls-and-protection-crisis-eastern-drc>.

[6] Lindsey (n 25).

[7] OHCHR, 'Gendered Impact of the Conflict in the Syrian Arab Republic on Women and Girls', (Web Page, 12 June 2023). <https://www.ohchr.org/en/statements-and-speeches/2023/06/gendered-impact-conflict-syrian-arab-republic-women-and-girls>.

to illness due to their special reproductive roles in childbirth, pregnancy and delivery, leading to a stronger need for health services.[1] In some countries caught in humanitarian crises, women face more barriers than men in accessing health care during armed conflict, especially in the area of sexual and reproductive health services, due to complex and diverse environments and large gender gaps. For instance, an estimated 140,000 women die in conflict each year, and war impedes access to health services, among which the widespread failure of sexual health services prevents women from accessing safe and effective contraception, abortion and treatment for sexually transmitted diseases.[2] Therefore, countries where gender discrimination exists in conflict can also limit women's active access to treatment and disease prevention. For example, women may be limited by cultural norms, some women may only travel long distances to seek medical care if accompanied by a male relative, and in some economically constrained communities, the limited resources available to pay for health care often go to the male family members rather than to women and children, especially when armed conflict has a disproportionate impact on people's normal lives.[3] These cultural and religious requirements can make it more difficult for women to receive appropriate health care when they need it. Women are also at high risk of contracting sexually transmitted diseases in conflict situations. This is mainly because they may be forced to have sex in the war due to the lack of resources and the increase in sexual violence, so that they can get food and water and other resources to ensure the basic life of themselves and their children. Sexually transmitted infections not only cause a series of serious traumas for women (such as ectopic pregnancy and infertility), but also have a high risk of transmission of sick pregnant women to their newborns.[4] Thus, causing more serious consequences. Finally, in addition to women's difficulties in accessing specific health services in situations of armed conflict, women's access to health services in general may face challenges based on religious requirements and gender discrimination, and they may be treated unfairly when requesting medical assistance (e.g. free health services still require payment), and women may be treated unfairly when requesting medical assistance; they may also be subject to abuse (e.g. rough treatment and neglect of treatment needs) by medical personnel who help them.[5] This is not only detrimental to women's physical health protection, but also may make them suffer psychological trauma and other negative effects.

Women are generally underrepresented in decision-making processes in the humanitarian field.[6] In 2016, women accounted for 42.8% of all employees in the UN system, with a higher concentration in entry-level positions, and only 9 out of 29 UN humanitarian coordinators were women, which not only means that the gender gap is widespread in the world today, but also has a potential impact on the implementation of gender protection in the humanitarian field.[7] For example, in armed conflicts, the work of many organizations is to provide targeted support to women and children who are more vulnerable in war through social support and psychological counseling, which can not only be understood as empowering women's rights, but also effectively treat the physical and mental trauma of victims of sexual violence, which has positive significance.[8] However, if women's participation in decision-making is ignored, it is not only conducive to a comprehensive understanding of the special needs of women receiving support, but also may hinder the direct feedback of women in these areas, resulting in the weakening of service delivery effect and the failure to achieve the expected goal of aid. Gender discrimination is considered a barrier to women's participation in decision-making processes, particularly those related to conflict. For example, in a study of women's participation in decision-making

[1] Juan Carlos Rivillas et al. (2018). How do We Reach the Girls and Women who are the Hardest to Reach? Inequitable Opportunities in Reproductive and Maternal Health Care Services in Armed Conflict and Forced Displacement Settings in Colombia. *PLoS One, 13*(1) e0188654, 1–14, 5.

[2] Achier D Akol, Silke Caluwaerts and Andrew D Weeks. (2016). Pregnant Women in War Zones. *BMJ, 353*, i2037, 1–2, 1.

[3] Lindsey (n 25) 111.

[4] Ibid 112.

[5] Lindsey (n 25) 113.

[6] Spoljaric (n 22).

[7] Ayla Black, Pip Henty and Kate Sutton. (2017). *Women in Humanitarian Leadership*. Humanitarian Advisory Group, 4.

[8] Lewis Turner. (2016). Are Syrian Men Vulnerable Too? Gendering The Syria Refugee Response. (Web Page, 29 November 2016) <https://www.mei.edu/publications/are-syrian-men-vulnerable-too-gendering-syria-refugee-response>.

processes in Colombia, while women's participation in regional discussion forums increased during the recent peace process, women's underrepresentation in peace negotiations during the ceasefire meant that gender discrimination due to women's marginalization in historical and cultural contexts continued to persist; women are not seen as important transformative forces in conflict.[1] Whether in conflict or post-conflict reconstruction, such stereotypes will not only lead to the continuation of gender inequality, but also fail to effectively obtain the true voice of women in conflict, thus unable to achieve the ultimate goal of gender protection.

In summary, women in armed conflict are significantly underrepresented in terms of access to resources, health-care services and representation in decision-making bodies, which in turn is rooted in the realities of sexual violence and gender discrimination that are prevalent in conflict. Due to the particular vulnerability of women, gender-specific protection planning under the IHL framework is essential not only to ensure comprehensive protection and relief for civilians in conflict, regardless of gender or ethnicity, to curb the negative effects of armed conflict, but also to improve the reconstruction of post-conflict mechanisms and promote the achievement of more far-reaching gender equality goals.

3. Insufficient Gender Protection in IHL Legal Provisions: Ambiguity and Barriers to Implementation

Although some provisions of the IHL have been discussed in the first part of this essay with regard to the regulation of gender-based violence and the protection of women's rights and interests, these provisions are still potentially imperfect and ambiguous, so that they fail to fully reflect the need for gender-specific protection in the context of armed conflict and the continuing gender differences. This is mainly reflected in the targeted criticism of some feminist legal theorists on the gender differences existing under the framework of IHL. The key point is that these scholars believe that IHL essentially advocates the priority of men, and only when women are victims and bear the responsibility of raising children, they are given a legitimate status by law.[2] To a certain extent, this indicates that IHL pays more attention to formal equality in gender protection rather than substantive justice when armed conflicts have different impacts on different genders.

Under the framework of the IHL, various conventions, in particular the Geneva Conventions, provide for the protection of civilians in the course of armed conflict, while emphasizing that women civilians are equally protected as men. For example, with regard to the treatment of protected persons, article 27 of the Fourth Geneva Convention provides that women shall be accorded special protection among protected persons against any form of violation of their honour, such as rape or forced prostitution.[3] Article 16 of this Convention also makes it clear that pregnant women need special protection and respect in armed conflict, along with the wounded and sick.[4] In addition, with regard to the distribution and transport of relief supplies, article 23 of this Convention requires that States parties to armed conflict should allow the free passage of essential food, clothing and supplements for children, pregnant women and women giving birth, and that priority should be given to such groups in the distribution of relief supplies.[5] Unfortunately, the IHL rules that exist to protect the rights of civilians and women sometimes have a limited impact. For example, article 27 of the Fourth Geneva Convention, referred to above, provides for special protection of women's honour from the perspective of women as victims of gender-based violence in war (as opposed to men), often framed in terms of women's need for chastity and modesty rather than directly addressing the real physical and psychological damage to women caused by

[1] Pilar Domingo and Veronica Hinestroza. (2013, September 1). *Evaluation of UN Women's Contribution to Increasing Women's Leadership and Participation in Peace and Security and Humanitarian Response: Headquarter and Global Case Study.* Research Report, 85.

[2] Helen Durham and Katie O'Byrne. (2010). The Dialogue of Difference: Gender Perspectives on International Humanitarian Law. *International Review of the Red Cross*, *92*(877), 31, 34.

[3] International Committee of the Red Cross, 'Article 27 - Treatment I. General observations', (Web Page, 1949). <https://ihl-databases.icrc.org/en/ihl-treaties/gciv-1949/article-27?activeTab=>.

[4] International Committee of the Red Cross, 'Article 16 - Wounded and sick I. General protection', (Web Page, 1949). <https://ihl-databases.icrc.org/en/ihl-treaties/gciv-1949/article-16?activeTab=>.

[5] International Committee of the Red Cross, 'Article 23 - Consignment of Medical Supplies, Food and Clothing', (Web Page, 1949). <https://ihl-databases.icrc.org/en/ihl-treaties/gciv-1949/article-23?activeTab=>.

sexual violence in conflict.[1] This wording can have a strong negative impact on women by creating a widespread perception that women who have already been sexually assaulted are dishonored. Although the original intention of this provision in the IHL was well-intentioned, it effectively allowed the drafters of the convention to determine the definition of women's honour, which could lead to the further entrenchment of stereotypes and antiquated notions that view women as victims.[2] Second, IHL's protection of women is not based on them as an independent individual, but is often viewed from the perspective of a woman's relationship with others (such as family or men). This is reflected in the Geneva Conventions and their Additional Protocols of 1977, which contain a total of 42 articles specifically referring to the protection of women, but almost half of the articles are based on the role of women as pregnant women or breastfeeding mothers.[3] This shows that even though there are provisions against sexual violence under IHL, women are still viewed as the 'property' of their husbands or families. When they suffer sexual violence and their honour is damaged, the honour of the men they are attached to is also considered to have been violated; the logic behind this idea is that in armed conflict, men wage war against men through the violation of women's bodies.[4] This highlights the interactive relationship between IHL and the gender norms prevailing in society. The legal provisions under the framework of IHL are both the product of social gender concepts, and in turn reinforce these inherent concepts. The existence of this relationship makes gender bias form a vicious circle in legal norms and general social life, and it is difficult to break it.

In addition to the protection of women in the civilian context, the Geneva Conventions contain several provisions specifically for women combatants, most of which relate to women prisoners of war. For example, it is mentioned in the first part of this essay that articles 25 and 29 of the Third Geneva Convention stipulate that female prisoners of war should have separate dormitories and health facilities. In addition, Article 12 of the Second Geneva Convention on the care of the wounded in the armed forces requires that women should receive adequate care corresponding to their sex without any discrimination.[5] These special protections for women combatants lack a degree of consideration, that is, the Convention's restrictive provisions for combatants in armed conflict prevent many women from being formally qualified as combatants, mainly because of their often irregular combat roles and the traditional gender bias that may exist in local armed groups.[6] For example, Article 1 of the Hague Convention of 1907 clearly states that militia and volunteer regiments have rights and obligations under the laws of war only if they comply with the requirement to carry arms openly, have a special emblem, and have a commander.[7] Additional Protocol I to the Geneva Conventions also defines the scope of combatants of a party to the conflict, that is, the organized armed forces, groups and units of a party to the conflict, with the exception of medical personnel and chaplains, under the command of the commander of that party.[8] These restrictions on combatant eligibility exclude some women in support and supporting roles in combat (such as medical support personnel and spies), as well as women in irregular combat who are unable to meet the strict conditions, from a range of protections under the article, which is one reason why the restrictive definition of the IHL law is biased against women.

In the framework of IHL, there exists a potential power hierarchy, that is, combatants take precedence over civilians, public interests take precedence over personal interests, and military interests are superior to other interests. These explicit hierarchies often implicitly express the privileges of men over women.[9] This means that

[1] Durham and O'Byrne (39) 35.

[2] Orly Maya Stern. (2018). Gender and International Humanitarian Law' in Orly Maya Stern (ed), *Gender, Conflict and International Humanitarian Law: A Critique of the 'Principle of Distinction'*. Routledge, 103.

[3] Durham and O'Byrne (39) 34.

[4] Stern (n 44) 103.

[5] International Committee of the Red Cross, 'Article 12 - Protection and care of the Wounded, Sick and Shipwrecked', (Web Page, 1949). <https://ihl-databases.icrc.org/en/ihl-treaties/gcii-1949/article-12?activeTab=>.

[6] Stern (n 44) 101.

[7] International Committee of the Red Cross, 'Regulations: Art. 1', (Web Page, 1907). <https://ihl-databases.icrc.org/en/ihl-treaties/hague-conv-iv-1907/regulations-art-1?activeTab=>.

[8] International Committee of the Red Cross, 'Article 43 - Armed Forces', (Web Page, 1977). <https://ihl-databases.icrc.org/en/ihl-treaties/api-1977/article-43>.

[9] Stern (n 44) 104.

the IHL is not an important component of the rules relating to women as opposed to advancing the interests of men, especially combatants. For example, rape is not listed as a "serious violation" of the rules of the Geneva Convention against crimes against women, which is intended to protect women rather than specifically; this is a striking indication that gender crimes do not rank highly in the rigid hierarchy of war crimes.[1] In recent years, the entire international community has emphasized that rape is an egregious violation of the personal dignity of women, and many efforts have been made to that end, such as the first conviction of rape as a crime against humanity by the International Criminal Tribunal for Rwanda.[2] This additional step shows that crimes against women were not taken seriously in the initial armed conflict. Moreover, in 2013, the Declaration on the Prevention of Sexual Violence in Conflict, adopted by the original G8 member states in response to sexual violence in armed conflict, recognized rape and other forms of serious violent crime as war crimes in serious violation of the Geneva Conventions.[3] Although this has certain positive significance for the prevention of sexual violence in conflict, a declaration limited to member states may not cover countries around the world where armed conflicts still exist, and the resolution and protection of women's rights and dignity in conflict still face challenges.

4. Inconsistencies and Lack of Uniform Standards in International Implementation

Because of the different contexts and situations of armed conflicts around the world, if the gender protection provisions of IHL are inconsistent in practice or lack of specific standards for implementation, their effect on curbing the occurrence of sexual violence in armed conflicts will be limited. Determining which conflict-related sexual violence violates IHL is therefore a prerequisite for harmonizing standards for the application of IHL treaties to States in conflict. Gender issues, in particular sexual violence, are diverse and widespread, as they can occur during situations of armed conflict and violence as well as in times of peace, and even sexual violence in armed conflict is not necessarily "conflict-related". IHL treaties do not actually include the term "conflict-related sexual violence" as an official legal term, and different institutions have different understandings of the definition, such as the United Nations, which describes the definition as being linked to conflicts and political disputes in time and geography, and may be reflected in the motives of the perpetrators, the characteristics of the victims, and the national atmosphere.[4] This means that IHL applies only to armed conflict or acts sufficiently linked to armed conflict, and not that all conflict-related sexual violence constitutes a violation of IHL and a war crime.[5] What kind of sexual violence is associated with armed conflict is often not easy to determine. Gloria Gaggioli put forward a case on this issue, that is, in the context of a non-international armed conflict, if a military commander rapes a soldier in an army barracks as a punitive measure, the act is not related to the armed conflict and therefore does not apply to IHL; But if, in the same circumstances, the commander raped a woman detained during an armed conflict, it would be a clear violation of IHL.[6] Some acts of sexual violence are not fully covered due to this ambiguity in determining the correlation between acts and armed conflicts, which reflects the possible inconsistency in the implementation of IHL in the actual application process. However, it is difficult to judge the precise criteria for the existence of links only through theoretical provisions and provide adequate countermeasures for diverse scenarios, and such judgments need to be continuously improved according to the accumulation of relevant cases.[7]

In addition, although the IHL provides relevant provisions on gender protection, the effectiveness of the specific implementation of the provisions largely depends on the

[1] Durham and O'Byrne (39) 35.

[2] Richard J Goldstone and Estelle A Dehon. (2003). Engendering Accountability: Gender Crimes Under International Criminal Law. *New England Journal of Public Policy, 19*(1), 121.

[3] Foreign & Commonwealth Office, 'Declaration on Preventing Sexual Violence in Conflict: Adopted in London on 11 April 2013', (Web Document, 11 April 2013), 2 <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/185008/G8_PSVI_Declaration_-_FINAL.pdf>.

[4] *Conflict-Related Sexual Violence: Report of the Secretary-General*, UN GAOR, 34th sess, Agenda Items 34, UN Doc A/66/657* and S/2012/33* (13 January 2012) 2.

[5] Gloria Gaggioli. (2014). Sexual Violence in Armed Conflicts: A Violation of International Humanitarian Law and Human Rights Law. *International Review of the Red Cross, 96*(894), 503, 514.

[6] Ibid 515.

[7] Ibid 517.

implementation and compliance of the parties to the conflict, and the differences in the system, ideology and culture of the armed groups as the subject of the conflict make it difficult to implement the consistency of the IHL provisions. Whether armed groups commit sexual violence against civilians can reflect ideological and institutional differences among different groups to a certain extent. For example, in the civil wars in El Salvador and Peru, various branches of the national army and some anti-government groups differ in the patterns of sexual violence (such as the frequency, scale and groups of victims of sexual violence), this corresponds to its different systems; At the same time, studies have shown that armed groups with mature and sustained political and ideological education mechanisms and rebel groups following a specific ideology (e.g., communism) are less likely to engage in sexual violence during armed conflict.[1] However, the IHL does not fully cover the internal factors of these complex armed groups, which may cause the general rules to be disconnected from the specific situations in the conflict, and thus the IHL lacks adequate coping mechanisms for similar phenomena.

The above phenomenon indicates that, as far as the content of the convention currently included in the IHL is concerned, there is a lack of an effective authority to unify the implementation and standards of the IHL. The Human Rights Council (HRC), established in 2006, is the body of the United Nations system responsible for the promotion and protection of human rights through its special procedural mechanisms and the monitoring and assessment of human rights phenomena around the world. [2] Therefore, could consideration be given to delegating some of the relevant powers provided for in the Convention to the HRC through specific provisions of the Convention, so as to achieve specific protection for groups suffering from gender issues? This consideration is mainly supported by the following two practical factors. First, the Human Rights Council has a universal periodic review (UPR) mechanism, under which Governments are required to submit reports every four and a half years on their efforts to protect human rights, which are reviewed and commented on by a working group composed of member States, based on information provided by the States concerned; the aim is to establish an objective and reliable cooperation mechanism based on active dialogue to improve the human rights situation of all countries, to promote the fulfilment of their human rights obligations and commitments and to share their experience in the protection of human rights, taking into account their level of development and specific circumstances. [3] Through regular monitoring and reporting mechanisms, not only can gender-based violence and the protection of women's rights in armed conflict be brought to wider attention and countries be encouraged to assess the effectiveness of relevant improvement measures. More flexible and targeted gender protection measures can also be achieved through the sharing of national experiences on how to deal with the issue and by taking into account the geographical and cultural specificities of each country. Second, the HRC established a Special Rapporteur on Violence against Women, its causes and consequences (SRVAW), within its special procedures; its mandate is to investigate gender-based violence against women through cooperation with governments, non-governmental organizations and relevant international organizations and to report thereon to the HRC. For example, gender-based crimes were explicitly included in the Statute of the International Criminal Court (ICC) in 2002, and this historic development is in line with SRVAW's recommendations in its report on violence against women in armed conflict, which also documents that violence against women in armed conflict remains high and persistent; it also points out that there is still a gap between the actual implementation of the relevant legal provisions and the implementation objectives.[4] This shows that the SRVAW's relevant reporting documents under the Human Rights Council are not only forward-looking in policy, but also can provide targeted assessment and supervision of the response to gender-based violence, thereby improving the transparency and consistency of the implementation of women-specific protection in countries. These existing

[1] Elisabeth Jean Wood. (2014). Conflict-Related Sexual Violence and the Policy Implications of Recent Research. *International Review of the Red Cross, 96*(894), 457, 469.

[2] Bertrand Ramcharan. (2011). *The UN Human Rights Council.* Routledge, 1.

[3] Ibid 50.

[4] Yakin Ertürk. (2018, December 19). *15 Years of the United Nations Special Rapporteur on Violence Against Women, Its Causes and Consequences.* Report, 17.

mechanisms can also provide relevant experiences and frameworks for IHL on gender protection issues to some extent.

Although it is theoretically feasible to delegate some of the powers of IHL to the HRC, there may still be challenges in the practical implementation process. For example, the HRC is mainly related to international Human Rights Law (IHRL), which applies not only during armed conflict, but also during peace, while the IHL is a legal framework specifically applicable in times of armed conflict, mainly for the purpose of protecting civilians, prisoners of war, the wounded, etc. Thus, how to involve effective mechanisms, such as special procedures of the HRC, in the monitoring and implementation of gender protection, without violating the existing structure of the IHL, involves coordination among complex legal systems. At the same time, since the members of the HRC are elected by the UN General Assembly and depend to a certain extent on the political will of the international community, the different political positions of different countries may hinder the realization of the mechanism. The realization of this vision will therefore require close evaluation and coordination to achieve the best results for gender protection in conflict, which will be a long-term process.

5. The inadequacy of IHL in the Punishment of Gender Crimes

Although the IHL has put forward many protective provisions in documents such as the Geneva Conventions that address gender issues in all aspects of armed conflict, such as article 27 of the Fourth Geneva Convention and Article 75 of the Protocol I additional to the Geneva Conventions of 1977, which explicitly prohibit sexual violence such as rape in armed conflict. However, under the IHL framework, there is no clear punishment mechanism for violators of relevant provisions. At present, the punishment of gender crimes in armed conflict is mainly based on the case law of the International Court of Justice and the ICC, as well as the adequacy of the domestic judicial system of the country concerned. The establishment of the International Criminal Tribunals for the former Yugoslavia and Rwanda has led the international community to recognize that sexual violence constitutes a war crime when it occurs in situations of armed conflict and to hold perpetrators criminally accountable.[1] For example, the trials of the relevant International Criminal Tribunals in the Kunarac, Akayesu and Čelebiči cases found that rape, like other acts, can constitute crimes against humanity, genocide and ill-treatment; Based on these cases, the ICC has also included rape, sexual slavery and other forms of sexual violence as war crimes in international or non-international armed conflicts.[2] These case laws will not only play a role in clarifying and developing the relevant gender protection laws, but also have a certain deterrent effect on the perpetrators. While these initiatives have some positive implications in promoting gender protection, a significant proportion of the International Criminal Court's pending cases related to sexual violence do not result in a conviction for the crime.[3] In some countries in armed conflict, while perpetrators of rape and other forms of sexual violence should be prosecuted and punished through accountability mechanisms and investigations, impunity for sexual violence still exists in many cases. For example, in the case study of sexual violence during conflict in Kenya, Liberia, Sierra Leone and Uganda, victims of sexual violence were reluctant to report their experiences of sexual violence or take other steps to seek protection, mainly due to the widespread perception that the current legal system in the country was inefficient and corrupt; As a result, survivors lack confidence and a sense of security that the legal system can effectively hold them accountable.[4] In addition, due to the general lack of infrastructure and public services in areas affected by armed conflict, there are significant obstacles to investigating and documenting sexual violence during conflict, which makes it difficult to collect evidence on related issues and promote subsequent prosecutions, thus preventing the legitimate rights and interests of victims of sexual violence from being protected.[5]

[1] Anne Marie De Brouwer et al. (2013). *Sexual Violence as an International Crime: Inderdisciplinary Approaches* (Intersentia) ch 3, 58.

[2] Vincent Bernard and Helen Durham. (2014). Sexual Violence in Armed Conflict: From Breaking the Silence to Breaking the Cycle. *International Review of the Red Cross, 96*(894), 427, 431.

[3] Ibid 432.

[4] Kim Thuy Seelinger. (2014). Domestic Accountability for Sexual Violence: The Potential of Specialized Units in Kenya, Liberia, Sierra Leone and Uganda. *International Review of the Red Cross, 96*(894), 539, 548.

[5] Ibid 550.

In view of the above, how the IHL reaches agreement on cooperation with the ICC and measures to address gender-based violence within countries is key to advancing the punishment of perpetrators of gender-based violence in armed conflict. To achieve this goal, prevention policies should be tailored to the causes of gender problems, and the effective implementation of relevant gender protection laws and regulations should be promoted in stages by analyzing the causes of gender problems, whether the relevant governance framework is real and effective, and the in-depth understanding of different conflict situations and people involved in gender-based violence.[1] Secondly, the effective implementation of the IHL's relevant laws and regulations depends to some extent on the efficiency of the relevant international tribunals and the respect of the armed forces for its establishment rules. Strengthening the trial efficiency of cases of sexual violence in conflict and carrying out training on gender protection in military and armed organizations with a high incidence of gender issues will help further improve the applicable standards of IHL and formulate more effective laws and regulations to regulate gender issues in conflict. Finally, one of the main difficulties is how the IHL's legal framework can be harmonized with national legal systems and how it can work with relevant organizations, such as ICC, to improve accountability mechanisms for perpetrators. It is believed that with the changes in the international situation and the progress of the worldwide human rights movement, IHL will be able to improve relevant laws and regulations in the future, so as to achieve effective protection for groups suffering from gender issues in conflicts.

6. Conclusion

To sum up, this essay mainly illustrates that although the IHL framework provides specific protection for gender issues in armed conflict, there are still ambiguities and deficiencies in terms of the details of the provisions and the response to gender differences in armed conflict. This is mainly reflected in the fact that IHL attaches more importance to formal equality between men and women in terms of gender issues, but does not adequately address the difficulties of women's access to resources and health care under armed conflict, and their low participation in decision-making bodies. Through the analysis of legal documents under the framework of IHL such as the Geneva Conventions, it can be seen that the provisions of IHL on addressing gender issues will differ in the actual implementation process due to various factors. In addition, the IHL does not have clear provisions for the punishment of gender-based crimes, particularly crimes of sexual violence, which makes accountability for such crimes difficult in situations of armed conflict.

Due to the complexity and diversity of gender issues in different countries and regions in armed conflict, it is difficult for this study to conduct a comprehensive and detailed assessment of the effectiveness of gender protection initiatives under the IHL framework. In addition, this essay focuses mainly on the gender protection of female groups in armed conflict, and does not discuss the sexual violence against men that has occurred in past conflict situations, which is therefore limited. In future studies on whether the IHL adequately addresses gender issues, further discussion could be made on how the IHL framework can improve its implementation mechanisms for gender protection, in particular by considering including organizations such as the HRC in the IHL's accountability mechanisms for gender-based violence. Finally, it is hoped that the IHL will not only provide more effective protection for women in armed conflict, but also contribute to the progress of gender equality in the international law system as a whole, through targeted systematic reforms on the accumulation of experience in addressing gender issues in armed conflict.

References

Articles/Books/Reports

Akol, Achier D, Silke Caluwaerts and Andrew D Weeks. (2016). Pregnant Women in War Zones. *BMJ*, *353*, i2037, 1–2.

Bernard, Vincent and Helen Durham. (2014). Sexual Violence in Armed Conflict: From Breaking the Silence to Breaking the Cycle. *International Review of the Red Cross, 96*(894), 427.

Black, Ayla, Pip Henty and Kate Sutton. (2017). *Women in Humanitarian Leadership*. Humanitarian Advisory Group.

De Brouwer, Anne Marie et al. (2013). *Sexual*

[1] Bernard and Durham (n 64) 433.

Violence as an International Crime: Inderdisciplinary Approaches. Intersentia.

Domingo, Pilar and Veronica Hinestroza. (2013, September 1). *Evaluation of UN Women's Contribution to Increasing Women's Leadership and Participation in Peace and Security and Humanitarian Response: Headquarter and Global Case Study*. Research Report.

Durham, Helen and Katie O'Byrne. (2010). The Dialogue of Difference: Gender Perspectives on International Humanitarian Law. *International Review of the Red Cross, 92*(877), 31.

Ertürk, Yakin. (2018, December 19). *15 Years of the United Nations Special Rapporteur on Violence Against Women, Its Causes and Consequences*. Report.

Gaggioli, Gloria. (2014). Sexual Violence in Armed Conflicts: A Violation of International Humanitarian Law and Human Rights Law. *International Review of the Red Cross, 96*(894), 503.

Goldstone, Richard J and Estelle A Dehon. (2003). Engendering Accountability: Gender Crimes Under International Criminal Law. *New England Journal of Public Policy, 19*(1), 121.

Henckaerts, Jean-Marie and Louise Doswald-Beck. (2005). *Customary International Humanitarian Law Volume I: Rules*. Cambridge University Press.

Inter-Agency Workshop on Integration of Gender into Needs Assessment and Planning of Humanitarian Assistance. (1999, July). Summary Report.

Inter-Parliamentary Union and UN Women. (2023). *Women in Politics: 2023*. Report.

International Committee of the Red Cross. (2022, June). *Gendered Impacts of Armed Conflict and Implications for the Application of International Humanitarian Law*. Report.

Kolb, Robert and Richard Hyde. (2008). *An Introduction to the International Law of Armed Conflicts*. Bloomsbury Publishing.

Ramcharan, Bertrand. (2011). *The UN Human Rights Council*. Routledge.

Rivillas, Juan Carlos et al. (2018). How do We Reach the Girls and Women who are the Hardest to Reach? Inequitable Opportunities in Reproductive and Maternal Health Care Services in Armed Conflict and Forced Displacement Settings in Colombia. *PLoS One*, *13*(1), e0188654, 1–14.

Seelinger, Kim Thuy. (2014). Domestic Accountability for Sexual Violence: The Potential of Specialized Units in Kenya, Liberia, Sierra Leone and Uganda. *International Review of the Red Cross*, *96*(894), 539.

Spoljaric, Mirjana. (2023, March 3). *Gender Equality and War: "No Humanity, Dignity, Peace Until International Humanitarian Law Is Upheld for All Genders"*. Speech, Columbia University Law School.

Stern, Orly Maya. (2018). "Gender and International Humanitarian Law" in Orly Maya Stern (ed), *Gender, Conflict and International Humanitarian Law: A Critique of the 'Principle of Distinction'*. Routledge.

United Nations High Commissioner for Refugees. (2014). *Woman Alone: The Fight for Survival by Syria's Refugee Women*. Report.

Wood, Elisabeth Jean. (2014). Conflict-Related Sexual Violence and the Policy Implications of Recent Research. *International Review of the Red Cross*, *96*(894), 457.

Treaties

Conflict-Related Sexual Violence: Report of the Secretary-General, UN GAOR, 34th sess, Agenda Items 34, UN Doc A/66/657* and S/2012/33* (13 January 2012).

Geneva Convention Relative to the Protection of Civilian Persons in Time of War, opened for signature 12 August 1949, 75 UNTS 287 (entered into force 21 October 1950).

Geneva Convention Relative to the Treatment of Prisoners of War, opened for signature 12 August 1949, 75 UNTS 135 (entered into force 21 October 1950).

Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I), signed 8 June 1977, 1125 UNTS 3 (entered into force 7 December 1978).

Other

Foreign & Commonwealth Office, 'Declaration on Preventing Sexual Violence in Conflict: Adopted in London on 11 April 2013', (Web Document, 11 April 2013). <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachm

ent_data/file/185008/G8_PSVI_Declaration_-_FINAL.pdf>

International Committee of the Red Cross, 'Article 12 - Protection and care of the Wounded, Sick and Shipwrecked', (Web Page, 1949). <https://ihl-databases.icrc.org/en/ihl-treaties/gcii-1949/article-12?activeTab=>

International Committee of the Red Cross, 'Article 16 - Wounded and sick I. General protection', (Web Page, 1949). <https://ihl-databases.icrc.org/en/ihl-treaties/gciv-1949/article-16?activeTab=>

International Committee of the Red Cross, 'Article 23 - Consignment of Medical Supplies, Food and Clothing', (Web Page, 1949). <https://ihl-databases.icrc.org/en/ihl-treaties/gciv-1949/article-23?activeTab=>

International Committee of the Red Cross, 'Article 27 - Treatment I. General observations', (Web Page, 1949). <https://ihl-databases.icrc.org/en/ihl-treaties/gciv-1949/article-27?activeTab=>

International Committee of the Red Cross, 'Article 43 - Armed Forces', (Web Page, 1977). <https://ihl-databases.icrc.org/en/ihl-treaties/api-1977/article-43>

International Committee of the Red Cross, 'Commentary of 1958: Article 38 - Non-repatriated persons I. General observations', (Web Page, 1958). <https://ihl-databases.icrc.org/en/ihl-treaties/gciv-1949/article-38/commentary/1958>

International Committee of the Red Cross, 'Commentary of 1960: Article 25 - Prisoners of War and Quarters', (Web Page, 1960). <https://ihl-databases.icrc.org/en/ihl-treaties/gciii-1949/article-25/commentary/1960?activeTab=>

International Committee of the Red Cross, 'Commentary of 2020: Article 14 - Respect for the Persons and Honour of Prisoners', (Web Page, 2020). <https://ihl-databases.icrc.org/en/ihl-treaties/gciii-1949/article-14/commentary/2020?activeTab=>

International Committee of the Red Cross, 'Commentary of 2020: Article 16 - Equality of Treatment of Prisoners', (Web Page, 2020) <https://ihl-databases.icrc.org/zh/ihl-treaties/gciii-1949/article-16/commentary/2020#12>

International Committee of the Red Cross, 'Commentary of 2020: Article 29 - Hygiene', (Web Page, 2020). <https://ihl-databases.icrc.org/en/ihl-treaties/gciii-1949/article-29/commentary/2020?activeTab=>

International Committee of the Red Cross, 'Commentary of 2020: Article 49 - General Observations on Labour of Prisoners of War', (Web Page, 2020). <https://ihl-databases.icrc.org/en/ihl-treaties/gciii-1949/article-49/commentary/2020?activeTab=>

International Committee of the Red Cross, 'Commentary of 2020: Article 108 - Execution of Judicial Penalties: Premises and Essential Safeguards', (Web Page, 2020). <https://ihl-databases.icrc.org/en/ihl-treaties/gciii-1949/article-108/commentary/2020>

International Committee of the Red Cross, 'Gender and IHL', (Web Page, 2024). <https://www.icrc.org/zh/law-and-policy/gender-and-ihl>

International Committee of the Red Cross, 'Regulations: Art. 1', (Web Page, 1907). <https://ihl-databases.icrc.org/en/ihl-treaties/hague-conv-iv-1907/regulations-art-1?activeTab=>

Lindsey, Charlotte, 'Women Facing War', (Web Document, October 2001). <https://www.icrc.org/sites/default/files/external/doc/en/assets/files/other/icrc_002_0798_women_facing_war.pdf>

OHCHR, 'Gendered Impact of the Conflict in the Syrian Arab Republic on Women and Girls', (Web Page, 12 June 2023). <https://www.ohchr.org/en/statements-and-speeches/2023/06/gendered-impact-conflict-syrian-arab-republic-women-and-girls>

Turner, Lewis, 'Are Syrian Men Vulnerable Too? Gendering The Syria Refugee Response', (Web Page, 29 November 2016). <https://www.mei.edu/publications/are-syrian-men-vulnerable-too-gendering-syria-refugee-response>

World Food Programme, 'Women, Girls and the Protection Crisis in Eastern DRC', (Web Page, 7 March 2024). <https://reliefweb.int/report/democratic-republic-congo/women-girls-and-protection-crisis-eastern-drc>

PIONEER
ACADEMIC PUBLISHING LIMITED

Studies in Law and Justice
ISSN 2958-0382
www.pioneerpublisher.com/slj
Volume 4 Number 5 October 2025

Federal Challenges in Implementing International Obligations

Tianqi Yu[1]

[1] The University of New South Wales, Sydney NSW 2033, Australia
Correspondence: Tianqi Yu, The University of New South Wales, Sydney NSW 2033, Australia.

doi:10.56397/SLJ.2025.10.05

Abstract

This paper analyzes the structural challenges federal governments face in implementing international treaty obligations by comparing two landmark cases: *Commonwealth v Tasmania* (1983) from the Australian High Court and *Vishakha v State of Rajasthan* (1997) from the Indian Supreme Court. Both cases reveal contradictions in power distribution within federal systems, but with different root causes: the *Tasmanian Dam* Case reflects jurisdictional conflicts between federal and state governments, while *Vishakha* exposes legislative inertia. Through judicial intervention, the two cases were resolved respectively by expanding federal powers and enacting judicial guidelines. The paper argues that while federalism creates coordination inefficiencies and implementation delays, these challenges can be overcome through improved constitutional frameworks, enhanced coordination mechanisms, and timely legislative action.

Keywords: federalism, implementation of international obligations, jurisdictional conflicts, external affairs power, legislative inertia, judicial intervention, *Commonwealth v Tasmania*, *Vishakha v State of Rajasthan*, world heritage convention, Convention on the Elimination of All Forms of Discrimination Against Women (CEDAW)

1. Introduction

The High Court decision in *Commonwealth v Tasmania*,[1] commonly known as the *Tasmanian Dam Case*, is a landmark in Australian constitutional law. In this case, we can notice the difficulties federal governments face in introducing legislation to implement international obligations, particularly in balancing federal and state powers. By comparison, the Indian Supreme Court's judgment in *Vishakha v State of Rajasthan*[2] reveals similar challenges in a different context. This paper argues that both cases underscore the inherent tension between federalism and the implementation of international obligations, and both reflect some of the drawbacks of federalism, albeit with different mechanisms of resolution.

2. Backgrounds

2.1 Commonwealth v Tasmania (1983)

The Tasmanian Dam Case was decided in 1983 by the High Court of Australia, marking a pivotal landmark in constitutional law. At first, the Tasmanian government planned to build a hydroelectric dam on the Franklin River to

[1] *Commonwealth v Tasmania* (1983) 158 CLR 1.

[2] *Vishakha v State of Rajasthan* AIR 1997 SC 3011.

promote local economic development. Due to its natural beauty and ecological diversity, this area was listed as a UNESCO World Heritage site. Thus, the Commonwealth claimed construction of the dam was prohibited under the *World Heritage Properties Conservation Act* 1975(Cth), which is also called the WHPC Act. However, the Tasmanian government argued that the WHPC Act was unconstitutional on the basis that it was beyond the powers of the commonwealth to legislate (Section 51 of the Constitution) and claimed that land and water management were state matters.

There exist two main issues, the first is external affairs power, whether the federal government can legislate in areas traditionally under state control by invoking its external affairs power to fulfill international treaty obligations. Since the Commonwealth does not have specific powers to legislate on the environment within its territory or the protection of national or world cultural heritage, it seeks to justify legislation by invoking various powers, including the power to legislate on external affairs, as well as the legislative power arising from the national nature of the Government in matters that are particularly appropriate for regulation at the national level.[1]

The second is jurisdiction conflicts between the federal and states, is land and water management belongs to the federal government or states. The Court split as to which particular sections of the Act and regulations were valid, but in the end, sufficient legislation survived to prevent the Tasmanian Hydroelectric Commission from proceeding with the construction of the dam without the permission of the responsible federal Minister.[2]

2.2 Vishakha v State of Rajasthan (1997)

This landmark case happened in a village in the state of Rajasthan where a female social worker Vishakha was brutally gang-raped.[3] Although India had ratified the Convention on the Elimination of All Forms of Discrimination Against Women (CEDAW) in 1993, no corresponding domestic legislation had been enacted to address sexual harassment in the workplace. [4] Thus the petitioners invoked CEDAW to argue for judicial intervention.

The case illustrates that India lacks a resolving mechanism to deal with such circumstances. As what petitioners appeal, it is essential for the Indian Government to legislate to give victims access to seek justice. The honorable after hearing all the facts of the case exercised the power of the Indian Constitution (U/A 32) and stated that not only crime has been committed against the women but also her fundamental rights had been infringed. The court further concluded by giving various guidelines in this respect that should be followed.[5]

There also exist two key issues. The first one is legislative gaps between domestic laws and international treaties. In India, until this case's judgment was given out, there was no law to govern this matter. [6] Although the Indian government has recognized CEDAW, it failed to translate international commitments into actionable legislation. The second one is overlapping jurisdiction between federal and state governments. Given India's federal structure, the issue involved in this case will cause jurisdictional disputes in the legislative process, which delay the legislative process.

3. Judgments and Challenges

3.1 Commonwealth v Tasmania (1983)

The High Court ruled in favor of the Commonwealth, with a 4-3 majority. In the judgment, the High Court of Australia held the view that "parts of the WHPC Act were constitutional because it implemented a treaty and the Commonwealth was empowered to do so under its external affairs power in Section 51 (xxix) of the Australian Constitution, therefore Tasmanian was not able to build the dam". The Court insisted that the external affairs power enabled the federal government to legislate to fulfill its international treaty obligations, even in areas traditionally under state jurisdiction. This expanded the scope of federal legislative authority, particularly concerning environmental protection and international

[1] Andrew C. Byrnes. (1985). The Implementation of Treaties in Australia after the Tasmanian Dams Case: The External Affairs Power and the Influence of Federalism. *Boston College International and Comparative Law Review*, *8*(2), 275, 293.

[2] Ibid 294.

[3] Saksham Chhabra. (2020). Critical Analysis on Sexual Harassment at Workplace of Women in India. *International Journal of Law Management &Humanities*, *3*, 1147, 1149.

[4] Palakdeep Kaur and Inderjeet Kaur. (2023). Sexual Harassment Laws: Its Analysis and Impact. *Indian Journal of Law and Legal Research*, *5*(1), 4.

[5] Saksham (n 6) 1149.

[6] Palakdeep (n 7) 6.

obligations. The reason why the Dam Case can be a landmark is that it expanded the federal government's power to implement international obligations, which set a precedent for the primacy of international treaties in domestic legislation.

According to the judgment, it is difficult to align state interests with federal obligations under international law, which leads to inefficiencies in the federal system. Due to federal-state jurisdiction disputes, the implementation of international commitments could be delayed or complicated. Nonetheless, one is forced to conclude that the federal system and the institutional and political difficulties that federalism engenders have been extremely important in retarding Australian participation in international treaty regimes.[1]

According to comparison, this may be due in part to the federal government's failure to put pressure on the states when they do not respond to its initiatives with a reasonable speed, but the presence of an extra layer of government creates organizational problems that don't arise in a unitary state. These problems have had a severely inhibiting effect on Australia's participation in international treaty regimes.[2]

3.2 Vishakha v State of Rajasthan (1997)

The Supreme Court, invoking CEDAW and Article 32 of the Indian Constitution, issued the *Vishakha Guideline*s, which served as a binding framework for addressing workplace harassment until legislation was enacted. The Court emphasized the significance of aligning domestic laws with international treaty obligations and effectively filled the legislative void.

The case exposed the government's inability to enact laws promptly to fulfill international commitments. However, the court stated that the UN Convention sets out various strict sexual harassment laws on an international scale, and claimed that to date, India has not enacted any strict sexual harassment laws, that there is no conflict between domestic and international legislatures, and that the international laws that India has ratified and the UN Conventions should apply, and ordered the implementation of these guidelines.[3]

The root cause of this case might be all the policies and regulations established in recent years have failed to implement.[4] For example, employers have been careless and implementing the same at the workplace which in turn hampers the safety of women.[5]

4. Similarities Between the Cases

According to the above analysis of the two cases, those two cases have three similarities.

The first one is both cases illustrate challenges to the federal government. In Australia, the jurisdictional conflicts between federal and states result in separating state interests from federal obligations under international law. In India, legislative inertia in the federal government causes the legislative gap between domestic laws and international treaties.

The second one is courts in both cases played a pivotal role in resolving federal inefficiencies by upholding the federal authority (*Commonwealth v Tasmania*) or enacting binding guidelines (*Vishakha*).

The third one is both cases rely on international treaties. Case *Commonwealth v Tasmania* used the World Heritage Convention and Case *Vishakha* used CEDAW, both treaties were used to justify the rationality of federal governments' actions.

5. Differences Between the Cases

Based on the analysis of the two cases, I have summarized three main differences between them. The first one is the root of the problem. In the *Dam Case*, the overlap in jurisdiction between the federal and states had led to a dispute over whether the state of Tasmania had the right to build the dam. However, in *Vishakha*, due to legislative inertia, there is a legislative gap in sexual harassment cases. Before the *Vishakha Guidelines* came into the picture, women had to take matters of Sexual Harassment at the Workplace by complaining under Sec 354 and 509 of the Indian Penal Code.

The second one is the type of challenge. In the *Dam Case*, the main challenge mainly came from resistance from states. Tasmanian government claimed the WHPC Act was unconstitutional, which the federal government used to limit the state's power. However, in *Vishakha,* the key challenge is legislative inertia.

The third one is the resolution mechanism. *Dam*

[1] Andrew (n 1)338.

[2] Ibid 339.

[3] Saksham (n 6) 1156.

[4] Ibid 1160.

[5] Ibid 1161.

Case solved the problem by expanding federal powers through external affairs. However, *Vishakha* solved the problem by enacting judicial guidelines to fill legislative gaps.

According to the above analysis, I insist that federal systems present structural challenges. Firstly, both cases demonstrate federal-state jurisdiction conflicts, which makes the implementation of international obligations delayed or complicated. Secondly, in federal systems, federal and state governments require coordination, resulting in inefficiencies. Lastly, due to systemic inefficiencies, courts must take actions to solve, such as mediate conflicts or fill legislative voids.

From my perspective, the challenges of federal systems can be mitigated through clearer constitutional provisions, better coordination mechanisms, and timely legislative action. Judicial intervention, as seen in both cases, provides a temporary solution but is not sustainable in the long term.

6. Conclusion

The *Commonwealth v Tasmania* and *Vishakha v State of Rajasthan* cases highlight distinct but interconnected challenges in federal systems. The former underscores jurisdictional conflicts between federal and state governments, while the latter emphasizes legislative inertia and judicial activism. Both cases illustrate the complexities of aligning domestic laws with international obligations in federal systems.

While these challenges are significant, they are not insurmountable. Federal systems can overcome these hurdles through enhanced coordination, clear constitutional frameworks, and proactive legislative action. Ultimately, these cases underscore the importance of balancing federal structures with the need for timely and effective implementation of international commitments.

References

Andrew C. Byrnes. (1985). The Implementation of Treaties in Australia after the Tasmanian Dams Case: The External Affairs Power and the Influence of Federalism. *Boston College International and Comparative Law Review*, *8*(2), 275, 293.

Palakdeep Kaur and Inderjeet Kaur. (2023). Sexual Harassment Laws: Its Analysis and Impact. *Indian Journal of Law and Legal Research*, *5*(1), 4.

Saksham Chhabra. (2020). Critical Analysis on Sexual Harassment at Workplace of Women in India. *International Journal of Law Management &Humanities*, *3*, 1147, 1149.

PIONEER
ACADEMIC PUBLISHING LIMITED

Studies in Law and Justice
ISSN 2958-0382
www.pioneerpublisher.com/slj
Volume 4 Number 5 October 2025

Implicit Gender Bias in Judicial Decision-Making in Domestic Violence Cases in South Africa

Ayanda Zulu[1]

[1] University of Limpopo, Polokwane, South Africa
Correspondence: Ayanda Zulu, University of Limpopo, Polokwane, South Africa.

doi:10.56397/SLJ.2025.10.06

Abstract

This study investigates implicit gender bias in judicial decision-making in South African domestic violence cases, focusing on how unconscious gendered assumptions shape reasoning, credibility assessments, and sentencing. Based on a qualitative analysis of 210 court judgments from Cape Town, Johannesburg, and Durban (2016–2022), along with 15 interviews with magistrates and prosecutors, the research identifies three recurring patterns of bias: *emotional credibility bias*, where women's emotions are viewed as unreliable; the *neutrality myth*, which disguises patriarchal reasoning as objectivity; and *reconciliation bias*, reflecting judicial preference for family harmony over accountability.

Findings show that these biases are reinforced by patriarchal norms, weak institutional training, and formalistic legal culture, which collectively undermine the equality principles of South Africa's *Constitution (1996)* and *Domestic Violence Amendment Act (2021)*. The paper argues that effective reform requires gender-sensitivity training, trauma-informed judicial practice, and stronger accountability mechanisms within the justice system. A shift toward gender-responsive justice—grounded in empathy, equity, and survivor-centered adjudication—is essential to realizing South Africa's vision of transformative justice.

Keywords: gender bias, judicial decision-making, domestic violence, implicit bias, feminist jurisprudence, South Africa, magistrates' courts, gender-based violence, transformative justice

1. Domestic Violence and Legal Context in South Africa

Domestic violence remains one of the most pervasive human rights challenges in South Africa, rooted in deep historical inequalities and perpetuated by structural gender norms. According to *Statistics South Africa's 2022 Crime Against Women Report*, approximately 45% of women have experienced physical, emotional, or sexual abuse by an intimate partner during their lifetime, while only about 25% of these incidents are reported to authorities. Despite a robust constitutional framework guaranteeing gender equality, the persistence of domestic and gender-based violence (GBV) reflects a troubling gap between legal protection and practical enforcement.

The legal foundation for addressing domestic violence in South Africa is primarily established through the *Prevention of Domestic Violence Act (PDVA) 116 of 1998*. This landmark legislation defines domestic violence broadly —

encompassing physical, sexual, emotional, verbal, psychological, and economic abuse — and seeks to provide accessible protection orders for victims. It represents a major shift from the apartheid-era legal system, which often dismissed domestic violence as a private matter. The Act emphasizes immediate judicial intervention through protection orders and police obligations to assist victims, signaling a transition toward a more rights-based and survivor-centered legal framework.

However, more than two decades after its enactment, implementation challenges continue to undermine the PDVA's objectives. The Department of Justice and Constitutional Development (DOJCD) (2021) reports that while the number of protection orders issued has steadily increased — reaching nearly 280,000 applications annually — enforcement remains inconsistent across regions. Many victims still face procedural delays, limited police responsiveness, and insufficient follow-up by magistrates. For example, in some rural jurisdictions, magistrates have been found to encourage "reconciliation" between parties rather than legal sanction, reflecting persistent cultural perceptions that prioritize family unity over women's safety.

The introduction of the Domestic Violence Amendment Act (2021) aimed to strengthen accountability mechanisms by mandating improved inter-agency coordination and the electronic tracking of protection orders. It also expanded definitions to include "controlling behavior" and "coercive control," aligning domestic law with global standards, including the UN Convention on the Elimination of All Forms of Discrimination Against Women (CEDAW). Nonetheless, successful implementation depends heavily on the judiciary's interpretation and application of these provisions. Judges and magistrates act as critical gatekeepers in determining whether victims receive adequate protection or encounter secondary victimization through dismissive or biased adjudication.

Judicial attitudes toward domestic violence cases have thus become a central concern in the broader discourse on gender justice in South Africa. Studies conducted by the *Commission for Gender Equality (2020)* and *Gender Links (2021)* reveal that a significant proportion of judicial officers hold ambivalent views about the seriousness of domestic violence, often influenced by gender stereotypes. For instance, some magistrates reportedly perceive domestic disputes as "mutual conflicts" rather than abuse, leading to mitigated sentencing or informal mediation approaches. This tendency underscores how implicit gender bias — even in the absence of overt discrimination — can distort judicial neutrality and weaken legal protection for victims.

Furthermore, the National Strategic Plan on Gender-Based Violence and Femicide (2020–2030) acknowledges judicial bias as a systemic barrier to justice. The plan calls for gender-sensitivity training for all judicial officers and enhanced monitoring of court outcomes in domestic violence cases. It also emphasizes that a justice system sensitive to gender dynamics is essential not only for deterrence but also for the restoration of public trust in the rule of law.

2. Nature of Implicit Gender Bias

Implicit gender bias refers to the unconscious attitudes, stereotypes, and mental shortcuts that influence decision-making without a judge's explicit awareness or intent to discriminate. Unlike overt prejudice or deliberate bias, implicit bias operates automatically — shaping how judicial officers interpret evidence, assess credibility, and apply the law in domestic violence cases. In the South African context, this phenomenon is particularly significant given the coexistence of a progressive constitutional commitment to gender equality and the persistence of patriarchal social structures that subtly influence cognition and perception.

Research in cognitive psychology and social neuroscience demonstrates that implicit bias arises from schema-based associations — learned social patterns that link certain traits or behaviors to specific groups. For example, deeply ingrained societal beliefs about gender roles may unconsciously lead judicial officers to perceive men as protectors and women as emotional or unreliable witnesses. In courtroom contexts, such associations can manifest in differential evaluation of testimony, where a male defendant's denial is unconsciously afforded more credibility than a female victim's account, even when objective evidence is comparable.

South African legal scholars and gender researchers have increasingly identified implicit bias as a hidden barrier to substantive equality

in the judiciary. A 2021 study by *Gender Links* involving interviews with magistrates across five provinces found that over 40% of respondents viewed domestic violence disputes as matters that "could be resolved privately," reflecting unconscious minimization of the harm involved. Similarly, the *Commission for Gender Equality (2020)* reported instances where judges used language implying that victims were partly responsible for the violence due to "provocation" or "failure to maintain family harmony." Such patterns do not necessarily stem from conscious misogyny but from the internalization of cultural narratives that normalize male authority and female accommodation.

Implicit gender bias also intersects with institutional culture and systemic norms. The South African judiciary, historically shaped by colonial and patriarchal influences, has long privileged notions of rationality and detachment often associated with masculinity. As feminist legal theorists such as Catharine MacKinnon and African scholars like Penelope Andrews have argued, the myth of judicial neutrality can obscure the gendered assumptions embedded in law itself. In this sense, implicit bias is not merely an individual cognitive flaw but a structural feature of legal reasoning, sustained through precedent, training, and professional socialization.

Moreover, implicit bias may operate through linguistic framing and procedural choices. For instance, magistrates might unintentionally use euphemistic or neutral language ("family conflict," "relationship issue") that downplays the violent nature of abuse. Studies of court transcripts have shown that such framing can influence case outcomes, making it less likely for domestic violence to be perceived as a criminal violation warranting full legal intervention. These subtle cognitive distortions, though unintentional, contribute to what scholars term secondary victimization — when victims experience further psychological harm through dismissive or biased judicial treatment.

The challenge of addressing implicit bias lies in its invisibility and resilience. Because it operates beneath conscious awareness, traditional judicial ethics frameworks — which prohibit overt discrimination — are insufficient to prevent it. Awareness training and reflective practice are therefore critical for enabling judges to recognize how social conditioning may unconsciously shape their reasoning. The *National Strategic Plan on Gender-Based Violence and Femicide (2020–2030)* explicitly identifies implicit bias as a priority area for judicial reform, calling for systematic integration of gender-sensitivity and bias-awareness programs in judicial education curricula.

3. Theoretical Perspectives

3.1 Social Cognition

The social cognition perspective provides a foundational framework for understanding how implicit gender bias shapes judicial decision-making in domestic violence cases. Rooted in cognitive and social psychology, social cognition theory emphasizes that human judgment and behavior are influenced by mental schemas—organized knowledge structures that individuals use to interpret information and make decisions. These schemas, shaped by social learning and cultural experience, enable efficiency in judgment but can also lead to systematic distortions when they involve stereotypical or biased associations.

Within judicial contexts, social cognition operates through automatic information processing. Judges and magistrates, like all decision-makers, rely on mental shortcuts when confronted with complex cases, limited time, and emotionally charged testimony. These shortcuts often draw upon culturally embedded gender norms—for example, the association of men with authority, rationality, or control, and of women with emotionality, dependency, or exaggeration. Such implicit associations can unconsciously influence how judges interpret conflicting accounts in domestic violence trials, even when they consciously believe they are being neutral.

Empirical research supports this mechanism. Psychological studies, such as those by Greenwald and Banaji (1995) on implicit social cognition, demonstrate that even individuals committed to egalitarian principles may display automatic biases measurable through tools like the Implicit Association Test (IAT). Applying this to legal settings, scholars such as Kang et al. (2012) argue that judicial officers' exposure to repeated cultural narratives—such as the idea that domestic conflicts are "private family matters"—can prime unconscious cognitive responses that favor reconciliation or underplay violence severity.

In the South African judicial environment,

where colonial and patriarchal histories have long shaped legal culture, these cognitive schemas are particularly entrenched. Historical legal discourse positioned men as heads of households and arbiters of discipline, framing domestic violence as a matter of social order rather than individual rights. Despite post-apartheid constitutional reforms emphasizing equality, residual schemas persist in the subconscious layers of judicial reasoning. For instance, a magistrate may unconsciously interpret a woman's reluctance to leave an abusive relationship as consent or emotional instability rather than a rational survival strategy shaped by economic dependency and fear.

Social cognition theory also explains how confirmation bias interacts with implicit gender schemas. Judges may selectively attend to evidence that aligns with pre-existing assumptions—such as interpreting a woman's emotional demeanor as exaggeration or viewing male calmness as credibility. These micro-level biases accumulate to create macro-level disparities in court outcomes, contributing to the systemic leniency often observed in domestic violence sentencing.

Furthermore, social cognition frameworks illuminate how institutional environments can reinforce or mitigate bias. Judicial norms emphasizing "objectivity" and "detachment" may inadvertently discourage reflection on personal bias, sustaining the illusion of neutrality. Conversely, environments that encourage bias recognition and cognitive self-awareness, such as reflective judicial education programs, can reduce automatic stereotyping. This supports the inclusion of bias-interruption mechanisms—like decision checklists and deliberation protocols—within judicial training curricula recommended by the *National Strategic Plan on Gender-Based Violence and Femicide (2020–2030)*.

3.2 Feminist Jurisprudence

The feminist jurisprudence perspective provides a critical theoretical foundation for analyzing how implicit gender bias operates within the judicial system, particularly in domestic violence adjudication. This framework challenges the notion that the law is neutral or objective, arguing instead that legal systems often reflect and reproduce patriarchal power structures. Feminist legal theorists contend that legal principles, courtroom procedures, and judicial reasoning have historically been shaped by male-centered assumptions, which marginalize women's experiences and constrain the interpretation of justice in gender-based violence cases.

Feminist jurisprudence emerged in response to the recognition that the "universal legal subject"—often assumed to be rational, autonomous, and dispassionate—is in practice modeled after male experience. This critique, articulated by scholars such as Catharine MacKinnon (1989) and Sandra Fredman (1997), reveals how ostensibly neutral legal doctrines can obscure gendered realities. For instance, the traditional emphasis on evidentiary objectivity and emotional restraint in court may discount the trauma responses of domestic violence survivors, who often present with emotional expression or inconsistency due to fear and psychological distress. These behaviors are then misread as exaggeration or unreliability, reinforcing systemic disbelief toward women's testimony.

In South Africa, feminist legal analysis aligns closely with the nation's constitutional commitment to substantive equality, as outlined in Sections 9 and 10 of the *Constitution of the Republic of South Africa (1996)*. However, as legal scholars such as Penelope Andrews (2001) and Shireen Hassim (2014) observe, substantive equality requires more than formal legal equality—it demands transformation of the social and institutional norms that perpetuate gender hierarchy. Courts, as sites of both legal interpretation and cultural reproduction, thus become arenas where patriarchal ideologies may either be challenged or reinforced.

Domestic violence cases reveal the persistence of what feminist theorists term "epistemic injustice"—the systematic devaluation of women's knowledge and experience. Judges may frame domestic violence as "relationship conflict" rather than as a violation of human rights, thereby minimizing harm and overlooking structural power imbalances. This reflects what feminist theorist Carol Smart (1989) describes as the "masculine voice of law"—a narrative that privileges rationality and control, often at the expense of empathy and contextual understanding. Within such a framework, women's suffering can be rendered legally invisible or morally ambiguous, especially when courts favor reconciliation over

protection.

Moreover, feminist jurisprudence illuminates how judicial bias is reinforced through institutional culture and professional norms. South Africa's legal education and judicial training have traditionally prioritized procedural correctness over reflective awareness of gender dynamics. As a result, many judges remain unaware of how their reasoning reproduces structural inequality. The *Commission for Gender Equality (2020)* has noted that even where laws are progressive, implementation falters when judicial officers lack training in gender sensitivity and trauma-informed adjudication.

Applying feminist jurisprudence to judicial practice therefore involves rethinking neutrality and impartiality. True impartiality, as feminist scholars argue, does not mean ignoring gender but recognizing how social power relations shape perception and judgment. In this view, acknowledging women's lived experiences and the context of coercive control enhances rather than undermines judicial fairness. South Africa's *National Strategic Plan on Gender-Based Violence and Femicide (2020–2030)* echoes this principle by emphasizing the need for gender-transformative justice—a system that actively challenges rather than passively accommodates patriarchal bias.

4. Judicial Patterns of Bias

4.1 Gendered Framing of Violence and Shared-Blame Narratives

One of the most pervasive manifestations of implicit gender bias in domestic violence adjudication in South Africa lies in the gendered framing of violence and the frequent use of shared-blame narratives. These interpretive patterns occur when judicial officers describe domestic violence incidents not as unilateral acts of abuse, but as mutual conflicts or relationship disputes for which both parties bear responsibility. This framing subtly diminishes the seriousness of the offense and undermines the victim's credibility, effectively transforming a human rights violation into a private disagreement.

Studies conducted by *Gender Links (2021)* and the *Commission for Gender Equality (2020)* indicate that such framing remains widespread in South African magistrates' courts. Judges and magistrates often refer to incidents as "marital discord," "domestic disputes," or "lover's quarrels," implying moral equivalence between the perpetrator and the victim. This linguistic framing reflects a cognitive bias rooted in patriarchal socialization, where domestic conflict is perceived as a normal, even expected, component of intimate relationships. By framing violence as an interpersonal disagreement rather than a structural expression of gendered power, the judicial narrative obscures patterns of control and coercion that define domestic abuse.

For instance, analysis of selected court judgments in the *Western Cape High Court* (2018–2021) reveals repeated instances where judges used neutral or conciliatory language to describe violent acts. In one judgment (State v. M, 2019), a magistrate referred to a husband's physical assault on his spouse as "a moment of temper in the context of marital stress," suggesting emotional provocation rather than criminal intent. Similarly, another case (S v. N, 2020) involved the dismissal of a protection order application after the magistrate concluded that "both parties contributed to the escalation of the situation." Such phrasing exemplifies the shared-blame narrative—a subtle but damaging interpretive bias that minimizes responsibility and perpetuates the idea that victims play a role in their own victimization.

This interpretive pattern is reinforced by broader societal norms that valorize family unity and stigmatize divorce or separation, especially in communities where economic dependence and cultural expectations constrain women's autonomy. Judicial officers, operating within this cultural milieu, may unconsciously prioritize reconciliation or preservation of the family unit over accountability and protection. A 2022 survey by *Statistics South Africa* found that over 35% of women who experienced domestic abuse did not seek legal help because they feared being blamed or not believed—an outcome directly linked to such judicial framing.

Moreover, shared-blame narratives often intersect with assumptions about female emotionality and male rationality, reinforcing traditional gender hierarchies. Female victims who display visible distress during testimony may be perceived as unreliable or manipulative, while calm male defendants are viewed as credible and composed. This dynamic aligns with what feminist legal theorists call the "credibility gap"—a systemic tendency to interpret women's emotional expression as exaggeration rather than evidence of trauma (Smart, 1989).

The implications of gendered framing extend beyond individual cases. They contribute to a normative bias within the judiciary, shaping future legal interpretations and influencing prosecutorial behavior. When courts consistently depict domestic violence as relational conflict, police officers and prosecutors may internalize similar biases, leading to undercharging or reluctance to pursue cases. This cyclical effect perpetuates institutional tolerance toward violence against women, despite South Africa's progressive statutory framework.

Recognizing and challenging this bias requires not only awareness training but also linguistic and conceptual reform within judicial discourse. As noted by the *National Strategic Plan on Gender-Based Violence and Femicide (2020–2030)*, the language used in judgments and proceedings plays a powerful role in shaping perceptions of justice. Reframing domestic violence as a violation of constitutional rights—rather than a matter of interpersonal tension—reasserts the moral and legal gravity of such offenses. It also aligns judicial practice with the state's constitutional obligation to uphold gender equality and protect the dignity of all persons.

4.2 Sentencing Disparities and Leniency Toward Male Offenders

Sentencing in domestic violence cases in South Africa frequently reflects implicit gender bias through patterns of leniency toward male offenders and the inconsistent application of punitive measures. Although the *Prevention of Domestic Violence Act (PDVA) 1998* and the *Domestic Violence Amendment Act (2021)* mandate strong legal protection for victims, judicial discretion in sentencing often reproduces gendered hierarchies that prioritize rehabilitation and family reconciliation over deterrence and justice.

Empirical analyses of South African case law reveal a disproportionate emphasis on male defendants' mitigating circumstances, such as emotional stress, intoxication, or family responsibilities, which are often cited to justify reduced sentences. In contrast, female offenders in domestic contexts tend to receive harsher treatment for comparable offenses—particularly when they deviate from traditional gender norms of submission or restraint. This disparity underscores how sentencing decisions are not purely legal calculations but are influenced by socially constructed notions of gender and morality.

A 2021 study by the *Centre for Applied Legal Studies (CALS)* at the University of the Witwatersrand examined 240 domestic violence sentencing judgments from magistrates' courts in Gauteng and the Western Cape between 2015 and 2020. The analysis found that 64% of male offenders received suspended or non-custodial sentences, while only 18% of cases resulted in imprisonment exceeding one year. In cases involving repeated assaults, courts often justified leniency by referencing the accused's role as a "family provider" or by citing the victim's alleged "provocative behavior." These rationales reflect implicit cognitive associations linking masculinity with authority and emotional instability, and femininity with compliance and moral responsibility.

One illustrative example appears in *State v. D (2018, Johannesburg Magistrates' Court)*, where the defendant—a man convicted of assaulting his wife—received a wholly suspended sentence. The magistrate reasoned that incarceration would "harm the family unit" and that the offender had shown "remorse." Conversely, in *State v. R (2019, KwaZulu-Natal)*, a woman convicted of retaliatory assault against her abusive partner received a two-year custodial sentence, with the court emphasizing her "failure to act within the expected limits of a spouse." These cases highlight how gendered expectations about behavior and emotion shape judicial reasoning and outcomes.

This leniency toward male offenders is further reinforced by the structural biases in plea bargaining and prosecutorial discretion. Research by *Gender Links (2021)* and the *South African Law Reform Commission (SALRC, 2020)* notes that prosecutors frequently negotiate reduced charges or plea deals in domestic violence cases to "ease court backlog," disproportionately benefiting male defendants. The cultural normalization of domestic violence as a private or "low-intensity" crime perpetuates judicial reluctance to impose severe sentences, despite clear statutory provisions.

Moreover, judicial reliance on reconciliation narratives—encouraging offenders to "work things out" with their partners—often undermines deterrence and reinforces patriarchal power dynamics. According to the *Department of Justice and Constitutional*

Development (2021), in nearly one-third of domestic violence cases, magistrates recommended counseling or mediation instead of formal punitive measures. While such alternatives can be appropriate in certain contexts, their overuse in gendered violence cases reveals a misplaced emphasis on relational repair over victim safety.

These sentencing disparities have profound implications for justice and deterrence. The leniency extended to male offenders not only erodes public confidence in the judicial system but also signals tolerance of gender-based violence at an institutional level. This dynamic contributes to the high rates of recidivism documented by *Stats SA (2022)*, which estimates that approximately 25% of convicted domestic violence offenders reoffend within two years. The data underscore how judicial leniency undermines the state's broader strategy to combat gender-based violence, as outlined in the *National Strategic Plan on GBV and Femicide (2020–2030)*.

Addressing these disparities requires both structural reform and cognitive retraining. Feminist legal theorists argue that sentencing must move beyond patriarchal logic that excuses male aggression as circumstantial or emotional. The judiciary should instead adopt a trauma-informed and equality-centered approach, one that considers the broader patterns of control and coercion inherent in domestic violence. As the *Commission for Gender Equality (2020)* recommends, judicial education should include case simulations and reflective exercises to identify and counter implicit bias in sentencing.

- Gendered framing of violence and shared-blame narratives
- Sentencing disparities and leniency toward male offenders
- Skepticism toward victim credibility and testimony

4.3 Skepticism Toward Victim Credibility and Testimony

One of the most enduring expressions of implicit gender bias within South Africa's domestic violence adjudication lies in the judicial skepticism toward victim credibility. Despite the progressive legal framework established by the *Prevention of Domestic Violence Act (1998)* and reinforced by the *Domestic Violence Amendment Act (2021)*, courtroom dynamics often reveal an undercurrent of doubt directed toward victims—particularly women—whose testimonies are frequently scrutinized for emotional consistency, perceived motive, and demeanor. This skepticism operates as a subtle yet powerful mechanism that undermines access to justice and perpetuates secondary victimization within the judicial process.

Empirical evidence from the *Commission for Gender Equality (CGE, 2020)* and *Gender Links (2021)* highlights that many magistrates and judges continue to rely on stereotypical assessments of victim behavior. In interviews conducted with judicial officers across five provinces, over 38% admitted that a complainant's emotional expression during testimony—such as crying or anger—made them question the accuracy of her account. Conversely, calm or restrained demeanor was often interpreted as a lack of trauma or insincerity. This evaluative bias reflects a deeply gendered double bind: women who display emotion are viewed as unstable or manipulative, while those who remain composed are perceived as untruthful or unaffected.

Case analysis further supports this pattern. In *State v. Mokoena (2018)*, the Gauteng High Court reduced an assault conviction on the grounds that the complainant's testimony appeared "overly dramatic" and "emotionally charged," despite corroborating medical evidence. Similarly, in *S v. Tsele (2020)*, a magistrate dismissed a domestic violence charge, citing inconsistencies in the victim's statements and her "failure to leave the abusive relationship" as indicative of fabrication. These judicial interpretations not only disregard the well-documented psychological impacts of trauma—such as fragmented memory and emotional volatility—but also mirror societal myths that hold victims responsible for their continued victimization.

The skepticism toward victims is compounded by the courtroom's evidentiary structure, which privileges rational, linear narratives over affective and contextual testimony. Feminist legal scholars, including Carol Smart (1989) and Elizabeth Schneider (2000), have long critiqued this epistemological bias, arguing that legal institutions treat women's experiences of violence as inherently suspect because they do not conform to the "male model" of credible evidence. In South Africa, where domestic

violence often occurs in private spaces without witnesses, judicial reliance on corroboration and consistency disproportionately disadvantages victims, particularly those from marginalized communities with limited legal representation.

Recent research by the *Centre for Applied Legal Studies (CALS, 2022)* reinforces this concern, finding that in over 50% of domestic violence cases reviewed, magistrates questioned the complainant's motives, often suggesting that claims were exaggerated for financial gain or custody advantage. This assumption aligns with patriarchal narratives that depict women as vindictive or manipulative, undermining their legitimacy as victims. The same study noted that only 28% of magistrates reported receiving formal training in trauma-informed adjudication, indicating a serious institutional gap in judicial preparation for gender-based violence cases.

Linguistic analysis of courtroom transcripts provides further insight into how skepticism is linguistically encoded. Judicial officers often employ mitigating or distancing language—phrases such as "alleged assault," "domestic disagreement," or "emotional tension"—that subtly delegitimize the victim's account. This rhetoric not only weakens the legal weight of testimony but also reinforces a cultural discourse that normalizes violence as part of intimate life. As the *National Strategic Plan on Gender-Based Violence and Femicide (2020–2030)* emphasizes, such patterns of judicial speech have real-world implications: they influence police behavior, prosecutorial decision-making, and public trust in the justice system.

Psychologically, this skepticism can retraumatize victims who already face significant emotional and social barriers to reporting abuse. Studies from *Stats SA (2022)* indicate that fewer than one in four victims of intimate partner violence pursue legal recourse, citing fear of disbelief or humiliation in court as a primary deterrent. When judicial officers question the credibility of those who do come forward, they not only perpetuate injustice but also signal to society that gender-based violence is negotiable within the boundaries of social tolerance.

To counteract these biases, feminist jurisprudence advocates for the integration of trauma-informed judicial practices, which recognize the psychological realities of abuse survivors. Such approaches emphasize empathy, contextual interpretation, and the rejection of stereotypical credibility tests. The *CGE (2021)* recommends judicial education programs that incorporate experiential learning, survivor testimony workshops, and reflective analysis to help judges recognize unconscious bias in credibility assessment. These reforms are essential to transforming courtroom culture from one of skepticism to one of dignity-centered adjudication.

5. Empirical Framework

5.1 Court and Case Selection from Major Urban Jurisdictions

This study's empirical framework is grounded in a multi-site qualitative research design focusing on domestic violence adjudication within South Africa's major urban judicial settings. The selection of courts and case materials was guided by the need to capture regional, demographic, and institutional diversity, while maintaining comparability in terms of caseload volume, jurisdictional authority, and access to court records. Urban jurisdictions were prioritized because they represent both the highest density of domestic violence filings and the most visible intersection of socio-legal reform and entrenched patriarchal attitudes.

The primary data sites include magistrates' courts located in Cape Town (Western Cape), Johannesburg (Gauteng), and Durban (KwaZulu-Natal)—three cities that collectively process approximately 40% of South Africa's domestic violence protection order applications each year, according to *Department of Justice and Constitutional Development (DOJCD, 2021)*. These jurisdictions were selected for three main reasons:

1) High Case Volume and Judicial Diversity

   Urban courts handle a wide range of domestic violence cases, from intimate partner assaults to complex coercive control scenarios, providing a comprehensive sample for identifying judicial patterns. Moreover, metropolitan areas employ magistrates from diverse legal and cultural backgrounds, which enables a broader analysis of how implicit gender bias manifests across different interpretive traditions.

2) Accessibility of Records and

Observability of Proceedings

Urban magistrates' courts maintain digitized case management systems, allowing for structured access to court transcripts, sentencing remarks, and protection order records under ethical research protocols. Cape Town and Johannesburg courts, in particular, have participated in previous judicial monitoring programs under the *Centre for Applied Legal Studies (CALS)* and *Gender Links*, providing a precedent for academic collaboration and transparency.

3) Relevance to National Policy and Reform

These cities are key implementation sites for the *National Strategic Plan on Gender-Based Violence and Femicide (2020–2030)*, which emphasizes improving judicial responses and monitoring case outcomes. As hubs for feminist legal advocacy and public interest litigation, they reflect both the progressive ambitions and practical challenges of gender justice reform in South Africa.

The case selection process employed a purposive sampling strategy targeting domestic violence cases adjudicated between 2016 and 2022. This timeframe corresponds to the post-implementation phase of the *Victims' Charter (2007)* and precedes the enactment of the *Domestic Violence Amendment Act (2021)*, offering a balanced view of continuity and change in judicial reasoning. Cases were drawn from three sources:

- Protection order applications under the *Prevention of Domestic Violence Act (PDVA, 1998)*.
- Criminal assault and harassment cases involving intimate partners.
- Appeal judgments from regional courts and the High Court addressing evidentiary or procedural issues related to domestic violence.

In total, 210 cases were reviewed across the three jurisdictions: 80 from Gauteng, 70 from the Western Cape, and 60 from KwaZulu-Natal. These cases were selected using a combination of random and criterion-based sampling to ensure variation in judicial officer, defendant gender, and case outcome. Supplementary data were collected through courtroom observation (20 hearings across three cities) and semi-structured interviews with 15 magistrates. Ethical clearance for this fieldwork was obtained through the *University of the Witwatersrand's Human Research Ethics Committee (HREC)*, ensuring participant anonymity and compliance with the *Protection of Personal Information Act (POPIA, 2013)*.

This regional design facilitates a comparative analysis of judicial culture and implicit bias across different urban centers. Preliminary analysis revealed subtle jurisdictional differences—for instance, Cape Town courts tended to emphasize procedural formality and mitigation, while Gauteng courts showed higher variability in sentencing justification, often linked to judicial discretion and personal interpretation of victim behavior. Durban courts exhibited the highest rate of mediation-based resolutions, reflecting cultural influences that prioritize familial reconciliation.

The deliberate focus on these urban jurisdictions does not suggest that rural courts are free from bias, but rather acknowledges the empirical constraints of access and record availability. Nonetheless, the urban sample captures the core dynamics of bias and adjudication that shape national patterns of domestic violence jurisprudence. This methodological foundation enables a deeper exploration of how implicit gender bias manifests through judicial discourse, evidentiary evaluation, and sentencing practices within the framework of South Africa's evolving legal response to gender-based violence.

5.2 Qualitative Content and Interview-Based Data Analysis

The analytical design of this study integrates qualitative content analysis of judicial documents with semi-structured interviews conducted with magistrates and legal practitioners. This mixed qualitative approach enables both textual and experiential insights into how implicit gender bias operates in the adjudication of domestic violence cases in South Africa. By combining the analysis of court judgments with the voices of judicial actors, the study seeks to uncover not only *what* patterns of bias exist, but also *how* they are rationalized and reproduced within everyday judicial practice.

Documentary Analysis

The first component involved a systematic content analysis of 210 domestic violence case records drawn from magistrates' courts in Cape Town, Johannesburg, and Durban between 2016

and 2022. Each case included a full judgment text, sentencing remarks (if applicable), and procedural documentation relating to protection order applications or criminal assault proceedings. The documents were coded thematically using NVivo 12 qualitative analysis software, allowing for a structured identification of recurring linguistic and conceptual markers of bias.

The coding framework was guided by existing literature on judicial discourse and gender bias (e.g., Smart, 1989; Kang et al., 2012), as well as localized insights from the *Centre for Applied Legal Studies (CALS, 2021)* and *Gender Links (2021)*. Key coding categories included:

- Victim credibility framing (e.g., "exaggerated," "provoked," "mutual conflict")
- Defendant mitigation language (e.g., "momentary lapse," "family man," "economic stress")
- Reconciliation and relational language (e.g., "restore harmony," "family unity," "emotional reconciliation")
- Evidentiary skepticism (references to inconsistency, delay in reporting, or emotionality)

Two independent coders analyzed the documents to ensure inter-coder reliability, which achieved an agreement rate of 87% after iterative calibration. Discrepancies were resolved through discussion and by refining category definitions. NVivo frequency queries and word-tree visualizations were then employed to trace patterns in judicial language use, revealing consistent tendencies across jurisdictions—particularly the use of gendered euphemisms and minimizing language in describing acts of violence.

For example, the word frequency analysis showed that terms such as *"conflict," "dispute,"* and *"argument"* appeared five times more frequently than direct references to *"violence"* or *"assault"* in the corpus. Similarly, words denoting victim emotion—*"crying," "hysterical," "angry"*—were often paired with negative evaluative language like *"unreliable"* or *"irrational."* These findings empirically substantiate the claim that linguistic bias, even when subtle, materially influences judicial perception and reasoning.

Interviews with Judicial Officers

To complement the textual analysis, 15 semi-structured interviews were conducted with magistrates and prosecutors (five each from Cape Town, Johannesburg, and Durban). Participants were selected based on their active involvement in domestic violence adjudication or prosecution. Each interview lasted between 45 and 70 minutes and was conducted confidentially under the ethical clearance protocols approved by the *University of the Witwatersrand Human Research Ethics Committee (HREC)*.

Interview questions were designed to elicit reflection on judicial reasoning, emotional engagement, and perceptions of fairness in domestic violence cases. Thematic areas included:

- Interpretations of "reasonable behavior" by victims and defendants.
- Views on reconciliation and sentencing appropriateness.
- Awareness and acknowledgment of unconscious bias.
- Experiences with gender-sensitivity or trauma-informed training.

Thematic analysis of interview transcripts revealed a complex mixture of progressive awareness and residual patriarchal assumptions. While some magistrates expressed conscious commitment to gender equality, others described domestic violence as "emotional overreaction" or "mutual anger." Several participants explicitly stated that they viewed "excessive emotion" from victims as a sign of exaggeration or dishonesty, demonstrating how deeply internalized gender schemas continue to shape courtroom interpretation.

Interestingly, magistrates who had participated in judicial education programs organized under the *National Strategic Plan on GBV and Femicide (2020–2030)* showed greater self-awareness of potential bias. They were more likely to acknowledge that "neutrality" might unconsciously favor dominant cultural assumptions. However, only 6 out of the 15 participants reported having received such training, underscoring the limited institutional reach of current judicial sensitization efforts.

Analytical Integration

The combined analysis of documentary and interview data allowed for triangulation, strengthening the validity of findings. Textual

evidence of biased language was directly compared with magistrates' explanations and self-perceptions. This method revealed a striking cognitive dissonance: while most judges denied harboring gender bias, their language choices in written judgments reproduced precisely the stereotypes they claimed to avoid. This gap between self-perception and practice reflects the implicit nature of bias, confirming social cognition theory's insight that biases persist even among individuals with egalitarian intentions.

Furthermore, the integration of qualitative data illuminated the institutional conditions that sustain bias—such as time pressure, emotional fatigue, and reliance on precedent—creating fertile ground for cognitive shortcuts. These findings reinforce the necessity for reflective judicial practices, where judges are trained to critically interrogate their own interpretive patterns rather than assume neutrality as a default position.

5.3 Thematic Coding to Identify Bias Indicators

The thematic analysis of court judgments and judicial interviews revealed a complex network of implicit biases embedded within the reasoning, language, and practices of domestic violence adjudication in South Africa. Using an inductive grounded theory approach, key indicators of bias were identified through a process of open and axial coding with NVivo 12, supported by inter-coder validation and triangulation across data sources. The analysis uncovered three interrelated forms of bias that consistently shape judicial decision-making: emotional credibility bias, neutrality myth, and reconciliation bias.

The first, emotional credibility bias, refers to a recurring judicial pattern of assessing victim reliability through emotional expression rather than factual evidence. Women's testimonies were often dismissed as exaggerated or inconsistent when accompanied by strong emotion, while male defendants' emotional displays were interpreted as remorse or stress. This asymmetrical interpretation appeared in approximately two-thirds of analyzed judgments and reflects a deeply ingrained gender stereotype linking rationality with truthfulness and emotion with manipulation. Such perceptions undermine the evidentiary value of trauma narratives, ignoring well-established psychological research showing that fragmented or affective testimony is a common response to abuse.

The second indicator, the neutrality myth, captures how claims of judicial impartiality can conceal underlying gendered assumptions. Many magistrates portrayed themselves as "objective" or "dispassionate," yet their written judgments frequently described domestic violence as a "family matter" or "mutual conflict." These expressions create the illusion of balance while diffusing responsibility and minimizing the perpetrator's accountability. Interviews confirmed that neutrality was often equated with emotional detachment, a mindset that inadvertently favors patriarchal norms by dismissing women's lived experiences as subjective or excessive.

The third major indicator, reconciliation bias, concerns the judiciary's preference for maintaining family harmony over ensuring justice and protection. Sentencing remarks and protection order rulings often emphasized forgiveness and restoration rather than deterrence or accountability. In roughly one-third of the cases reviewed, courts explicitly encouraged reconciliation between the victim and the perpetrator, sometimes even suspending sentences to preserve family unity. This inclination reflects cultural norms that prioritize relational stability and male authority, positioning women's safety as secondary to familial cohesion.

When analyzed together, these biases reveal a systemic logic that transforms discretion into discrimination. Emotional credibility bias shapes how testimony is heard, neutrality myth distorts how evidence is weighed, and reconciliation bias influences how judgment is rendered. The interaction of these factors creates a cognitive and cultural environment where gender inequality is reproduced through ostensibly neutral legal processes. Even judges who explicitly reject sexism may unconsciously rely on these interpretive shortcuts due to institutional pressures, lack of trauma training, and inherited courtroom culture.

To ensure reliability, inter-coder agreement was established at 0.84 using Cohen's Kappa, and all coding categories were cross-validated with reference to external reports from the *Commission for Gender Equality (2020)* and *Gender Links (2021)*. Peer reviewers from the *University of Cape Town Centre for Law and Society* confirmed

that the identified themes aligned with broader research on judicial bias in gender-based violence cases. These results suggest that implicit bias within South African magistrates' courts is not isolated to individual judges but reflects broader institutional and cultural patterns embedded in the justice system.

Ultimately, the thematic coding confirms that judicial bias in domestic violence cases operates less through overt discrimination than through patterned assumptions about credibility, neutrality, and family reconciliation. The persistence of these cognitive frameworks highlights the need for structural reform and ongoing judicial sensitization to ensure that adjudication aligns with South Africa's constitutional commitments to gender equality, dignity, and justice.

6. Structural and Cultural Factors

The persistence of implicit gender bias in South African domestic violence adjudication cannot be understood solely at the level of individual cognition; it is sustained by broader structural and cultural forces that shape judicial behavior, institutional priorities, and societal expectations. Patriarchal norms, historical inequalities, and weak institutional accountability together form a cultural ecosystem that normalizes gendered interpretations of justice. In this context, judicial decisions become both a reflection of and a reinforcement for the social hierarchies embedded within South African society.

Patriarchal ideology remains a defining feature of the justice system's interpretive framework. Despite formal commitments to gender equality under the *Constitution of the Republic of South Africa (1996)* and the *Domestic Violence Amendment Act (2021)*, many judicial officers continue to operate within a worldview that associates male authority with order and female submission with respectability. This cultural inheritance, deeply rooted in colonial legal traditions and post-apartheid social hierarchies, shapes the implicit assumptions judges bring to the courtroom. Research by the *Commission for Gender Equality (2020)* and *Gender Links (2021)* indicates that judges often frame domestic violence as a private dispute rather than a structural manifestation of inequality, thereby reproducing the patriarchal logic that sustains gender-based violence.

The institutional structure of the South African judiciary also contributes to the endurance of bias. Lower magistrates' courts, where most domestic violence cases are heard, often operate under significant caseload pressure, limited resources, and inadequate training. Many judicial officers receive minimal or irregular exposure to gender-sensitivity or trauma-informed practice workshops, which are not mandatory across jurisdictions. The *Department of Justice and Constitutional Development (2021)* reports that only 27% of magistrates had completed certified training in gender-based violence adjudication by 2022, and even fewer had access to follow-up evaluation or peer mentoring. Without systematic training and reflection, judges are left to rely on their personal beliefs and social experiences, making them more susceptible to unconscious bias in interpreting victim behavior or evaluating evidence.

Culturally, South Africa's social landscape reinforces judicial conservatism in domestic violence cases. Traditional gender norms remain deeply entrenched in many communities, where masculinity is equated with authority, control, and emotional restraint. These values subtly infiltrate legal reasoning, leading some magistrates to interpret male aggression as a lapse in self-control rather than a deliberate act of coercion. Simultaneously, cultural expectations of female endurance and family preservation encourage judicial leniency and reconciliation-oriented outcomes. The *South African Law Reform Commission (2020)* found that reconciliation and mediation are disproportionately promoted in domestic violence cases compared to other violent crimes, reflecting an enduring belief that familial harmony outweighs individual safety.

Institutional culture further amplifies these patterns. The hierarchical nature of the judiciary discourages open discussion of implicit bias, as junior magistrates often emulate the reasoning and language of senior judges to align with prevailing norms. This socialization effect perpetuates biased interpretive habits across generations of judicial officers. Moreover, the formalistic approach to law—prioritizing procedural correctness over substantive justice—creates an environment in which discriminatory reasoning can persist without being recognized as such. Feminist legal theorists argue that this culture of formal neutrality masks systemic inequities by equating objectivity with detachment, effectively silencing

the gendered dimensions of domestic violence.

The political and institutional response to gender-based violence has also struggled to translate policy into practice. While the *National Strategic Plan on Gender-Based Violence and Femicide (2020–2030)* outlines a comprehensive reform agenda emphasizing survivor-centered justice, its implementation remains uneven. Budgetary constraints, fragmented coordination between the police, prosecutors, and courts, and limited data monitoring undermine the plan's transformative potential. As a result, judicial accountability for biased reasoning or inconsistent sentencing is rare. Complaints mechanisms, such as those managed by the *Judicial Service Commission*, tend to focus on procedural misconduct rather than interpretive bias, leaving cognitive and cultural prejudice unaddressed.

Importantly, cultural and institutional biases are mutually reinforcing. The judicial normalization of gender stereotypes validates public skepticism toward women who report abuse, discouraging victims from seeking legal recourse and thereby perpetuating the cycle of silence and impunity. This dynamic is particularly acute in marginalized communities, where socio-economic vulnerability intersects with race and class to amplify barriers to justice. As *Stats SA (2022)* notes, fewer than one in four women subjected to intimate partner violence pursue formal legal remedies, citing distrust in police and courts as a primary deterrent. This attrition not only weakens the legitimacy of the justice system but also entrenches the perception that domestic violence is a private issue beyond the scope of state intervention.

Addressing these structural and cultural factors requires a dual strategy of institutional reform and cultural transformation. On one hand, gender-sensitivity and trauma-informed training must be standardized and integrated into judicial education, accompanied by regular evaluation and peer review. On the other, broader public education campaigns are necessary to dismantle patriarchal narratives that normalize male authority and female endurance. Efforts by civil society organizations, such as *Sonke Gender Justice* and *Mosaic Training Services*, demonstrate that cultural change is achievable when legal reform is coupled with community-level engagement.

7. Reducing Judicial Bias

Reducing implicit gender bias in domestic violence adjudication requires a coordinated strategy that targets both cognitive awareness and institutional reform. While legal frameworks such as the *Domestic Violence Amendment Act (2021)* provide the statutory foundation for gender equality, the challenge lies in transforming judicial culture and practice. Bias cannot be eliminated solely through legislation—it must be confronted through deliberate, reflective, and systemic interventions within the judiciary itself.

A key starting point is the institutionalization of gender-sensitivity and trauma-informed training for all judicial officers. Current training programs in South Africa, though present, remain voluntary and sporadic. Data from the *Department of Justice and Constitutional Development (2021)* reveal that fewer than one-third of magistrates have completed certified programs on gender-based violence adjudication. Effective reform would require integrating bias awareness, social cognition, and trauma psychology into judicial education curricula, emphasizing how unconscious assumptions influence evidentiary evaluation and sentencing. Regular refresher workshops, peer discussions, and experiential learning modules—such as case simulations and survivor narrative analyses—could strengthen cognitive recognition of bias in real time.

Beyond individual awareness, structural mechanisms for accountability and reflection must be established. The *Judicial Service Commission* and *South African Judicial Education Institute (SAJEI)* should implement periodic reviews of judicial performance in gender-based violence cases, focusing not only on procedural compliance but also on interpretive fairness. Introducing a confidential judicial feedback and mentoring system, where judgments are peer-reviewed for language use and reasoning quality, would encourage reflection without punitive overtones. International examples, such as Canada's *National Judicial Institute* programs on implicit bias, demonstrate that consistent peer engagement can lead to measurable improvements in judicial reasoning and survivor treatment.

Institutional reform must also extend to data transparency and performance monitoring. Currently, domestic violence judgments are rarely published in accessible databases, limiting opportunities for external analysis and

accountability. Developing a centralized digital repository of anonymized case summaries could allow scholars, advocacy groups, and policymakers to track patterns of bias and evaluate progress over time. The *Commission for Gender Equality (CGE)* and *CALS* could collaborate on annual judicial review reports assessing trends in sentencing, language, and outcomes, providing an empirical foundation for continuous improvement.

Equally important is the integration of interdisciplinary expertise within the justice system. Collaboration with psychologists, social workers, and gender specialists can provide courts with holistic perspectives on victim behavior and trauma responses, reducing reliance on stereotypical interpretations. Specialized domestic violence courts—such as those piloted in Cape Town and Johannesburg—have shown promising results by combining legal adjudication with social support services, leading to higher victim satisfaction and improved case outcomes (Gender Links, 2021). Expanding these models nationwide would operationalize gender sensitivity within judicial structures.

Finally, judicial reform must be supported by cultural transformation within the legal profession. Law schools and professional associations should embed gender justice as a core component of ethics and judicial reasoning. Encouraging critical reflection on the intersection of law, power, and gender will help cultivate future magistrates who view fairness not as detachment, but as active equity. Sustained transformation depends on shifting the collective mindset of the judiciary from passive neutrality to conscious justice.

Reducing judicial bias, therefore, is both a technical and moral project. It demands institutional courage to confront tradition, self-awareness to challenge unconscious bias, and structural commitment to ensure that justice for victims of domestic violence is not undermined by the very system designed to protect them.

8. Toward Gender-Responsive Justice

Building a truly gender-responsive justice system in South Africa requires moving beyond procedural equality toward substantive justice that centers lived experience, empathy, and accountability. A gender-responsive judiciary recognizes that neutrality alone cannot ensure fairness when social structures remain unequal. Instead, it seeks to transform the conditions of adjudication so that survivors' voices are heard, validated, and protected within the legal process.

Central to this transformation is the redefinition of judicial objectivity. Traditional legal culture often equates objectivity with detachment, but a gender-responsive approach reimagines it as conscious engagement—an awareness of social context and power relations that inform every aspect of legal reasoning. This perspective aligns with South Africa's constitutional vision of *transformative constitutionalism*, which mandates not only the removal of discrimination but the proactive realization of equality in practice.

To achieve this, the justice system must be restructured around three guiding principles: empathy, accountability, and accessibility. Empathy demands that judges and prosecutors understand the psychological and social realities of gender-based violence, viewing victims not as unreliable witnesses but as participants navigating trauma. Accountability ensures that institutional practices, from police reporting to sentencing, are measured against gender-sensitive standards. Accessibility requires that victims—especially those from marginalized communities—can seek justice without intimidation, economic burden, or social stigma.

At the policy level, gender-responsive justice necessitates closer collaboration between the judiciary, civil society, and the executive. Initiatives such as the *National Strategic Plan on Gender-Based Violence and Femicide (2020–2030)* provide a roadmap for intersectoral coordination, but their success depends on consistent implementation and funding. Partnerships with advocacy organizations like *Sonke Gender Justice* and *Mosaic Training Services* should be institutionalized to ensure survivor-centered practices become the norm rather than the exception.

Cultural transformation remains the ultimate challenge. Legal reform alone cannot dismantle centuries of patriarchal conditioning that shape how society perceives domestic violence. Public education campaigns, community dialogues, and media engagement are essential to shift the narrative from tolerance to accountability. When the judiciary models empathy and fairness, it not only delivers justice within the courtroom

but also reshapes public expectations of justice itself.

References

Adrews, P. (2001). From gender apartheid to non-sexism: The pursuit of women's rights in South Africa. *North Carolina Journal of International Law and Commercial Regulation, 26*(3), 693–722.

Breen, T. D., & Wilson, T. D. (1994). Mental contamination and mental correction: Unwanted influences on judgments and evaluations. *Psychological Bulletin, 116*(1), 117–142.

Centre for Applied Legal Studies (CALS). (2021). *Sentencing patterns in domestic violence cases in South African magistrates' courts.* Johannesburg: University of the Witwatersrand.

Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis.* Thousand Oaks, CA: Sage.

Commission for Gender Equality (CGE). (2020). *Gender equality and the justice system: Annual review.* Johannesburg: CGE.

Constitution of the Republic of South Africa, Act 108 of 1996. Pretoria: Government of South Africa.

Department of Justice and Constitutional Development (DOJCD). (2021). *Annual performance report 2021/22.* Pretoria: DOJCD.

Domestic Violence Amendment Act 14 of 2021. Government Gazette, Republic of South Africa.

Fredman, S. (1997). *Women and the law.* Oxford: Oxford University Press.

Gender Links. (2021). *Judicial attitudes toward gender-based violence in Southern Africa.* Johannesburg: Gender Links.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review, 102*(1), 4–27.

Hassim, S. (2014). Voices, hierarchies, and the feminist politics of representation in South Africa. *Politikon, 41*(2), 193–209.

Kang, J., Bennett, M., Carbado, D., Casey, P., Dasgupta, N., Faigman, D., & Mnookin, J. (2012). Implicit bias in the courtroom. *UCLA Law Review, 59*(5), 1124–1186.

MacKinnon, C. A. (1989). *Toward a feminist theory of the state.* Cambridge, MA: Harvard University Press.

National Strategic Plan on Gender-Based Violence and Femicide (2020–2030). Pretoria: Government of South Africa.

Prevention of Domestic Violence Act 116 of 1998. Government Gazette, Republic of South Africa.

Protection of Personal Information Act 4 of 2013. Pretoria: Government of South Africa.

Smart, C. (1989). *Feminism and the power of law.* London: Routledge.

Sonke Gender Justice. (2022). *Transforming masculinities: Engaging men for gender justice in South Africa.* Cape Town: Sonke.

South African Law Reform Commission (SALRC). (2020). *Project 135: Sentencing and domestic violence.* Pretoria: SALRC.

Statistics South Africa. (2022). *Crime against women in South Africa report 2022.* Pretoria: Stats SA.

University of Cape Town Centre for Law and Society. (2022). *Peer review report on gender bias research methodologies.* Cape Town: UCT.

University of the Witwatersrand Human Research Ethics Committee. (2021). *Ethical guidelines for socio-legal research.* Johannesburg: Wits.

PIONEER
ACADEMIC PUBLISHING LIMITED

Studies in Law and Justice
ISSN 2958-0382
www.pioneerpublisher.com/slj
Volume 4 Number 5 October 2025

Conflict and Coordination Between Data Sovereignty and Digital Trade Liberalisation: An Analysis of Cross-Border Data Flow Policies in China and the EU

Xun Zhu[1]

[1] National University of Singapore, Singapore
Correspondence: Xun Zhu, National University of Singapore, Singapore.

doi:10.56397/SLJ.2025.10.07

Abstract

The digital economy faces a fundamental contradiction between the principles of digital trade liberalisation and the emerging demands of data sovereignty. This study analyzes this tension by conducting a comparative examination of the data governance frameworks of China and the European Union (EU). It contrasts China's state-centric approach with the EU's human rights-oriented model. The analysis demonstrates how these divergent models create significant non-tariff barriers to digital trade, clashing with the international liberalisation principles. It proposes several pathways for coordination to mitigate this tension. These include philosophical alignment on global justice, perfecting international trade rules with clear security exceptions, and strengthening regulatory cooperation.

Keywords: data sovereignty, digital trade liberalisation, cross-border data flow, international trade law

1. Introduction

In today's worldwide economic landscape, digital commerce and data transfer have brought about a radical transformation in the essence and extent of international business transactions. As data becomes a strategic resource, cross-border data flow supports all aspects of modern commerce, from supply chain management to digital service provision. [1] However, this digital transformation has created a fundamental contradiction between the principle of trade liberalisation and the idea of data sovereignty.[2] The "pre-internet" consensus, as represented by the agreements of the World Trade Organization (WTO), centered around facilitating the unhindered movement of goods and services across national borders.[3] Today, however, stakeholders in both developed and developing nations increasingly advocate

[1] Lateef MA. (2025). Digital Sovereignty in Global Trade: Analysing WTO Governance of Data Flows. *Beijing Law Review*, *16*, 875.

[2] Gao HS. (2021). Data Sovereignty and Trade Agreements: Three Digital Kingdoms. *SSRN Electronic Journal*.

[3] Appleton B. (2025). Digital Sovereignty vs. Trade Liberalization: India's Algorithm Disclosure Dilemma. Balsillie Case Studies. <https://balsilliecases.ca/case-study/digital-sovereignty-vs-trade-liberalization-indias-algorithm-disclosure-dilemma/> accessed 20 October 2025.

sovereign control of the data generated within their territories, citing concerns over national security, economic policy and citizens' privacy.[1] This has resulted in a "fragmentation of data privacy laws", creating legal uncertainty and threatening the integrity of the worldwide digital economy.[2]

At the heart of the conflict lies the matter of how cross-border data movements ought to be regulated within the context of international trade regulations. This study aims to tackle this issue by conducting a comparative examination of the data management systems in China and the EU. First, it establishes the theoretical foundations by contrasting the notion of data sovereignty with the principles of trade liberalisation. Second, it conducts a comparative analysis of Chinese and EU legal policies, identifying specific points of the conflict. Finally, it explores potential pathways for coordination, with the aim of establishing a sustainable equilibrium between the justifiable authority to regulate and the necessity of open digital commerce.

2. Theoretical Foundations of Data Sovereignty and Digital Trade Liberalisation

2.1 Data Sovereignty: Concept, Models, and Developments in the Digital Age

Sovereignty serves as a cornerstone of international law and a fundamental principle guiding international relations. Jean Bodin, the 16th-century thinker, was among the first to articulate a systematic theory of sovereignty, defining it as the "absolute and perpetual power of a commonwealth", a supreme authority unrestrained by the laws it creates.[3] The Dutch jurist Hugo Grotius further developed the international dimensions of sovereignty, analysing its internal and external perspectives. Internally, sovereignty denotes a state's right to control persons, events and things within its territorial boundaries. Externally, it denotes a state's right to be free from interference by other sovereign states.[4]

The advent of the digital age has extended this principle into a new domain. Data sovereignty and cyberspace sovereignty are now considered developments in sovereignty theory. Scholars generally consider data sovereignty as originating from cyber sovereignty, viewing it as a subset thereof.[5]

While data itself is intangible, it possesses certain physical attributes. Network data storage and infrastructure are typically located within a country's borders, and storage devices are usually owned by the state or corporations. Data transmission relies on national infrastructure such as power grids and cables. [6] The implementation of the idea of state sovereignty to data includes data storage devices within a nation's territory, which may be located in its territorial waters, land, or airspace. Under this framework, unauthorized access to data and its storage infrastructure by a foreign state is construed as a violation of the host state's sovereignty.[7]

2.2 Digital Trade Liberalisation: Core Principles Under the GATS Framework

Under the World Trade Organization (WTO), the General Agreement on Trade in Services (GATS)[8] has been set up. It serves as the main multilateral tool for regulating service trade, which encompasses digital or cross-border services. Its objective is to build a dependable and foreseeable framework of regulations for international service trade and to promote its gradual liberalisation.

The GATS places general duties on member states that are critical for digital trade. Article II establishes the principle of Most-Favoured-Nation Treatment. [9] This principle mandates that the WTO members should not show discrimination among their trading partners. Article XVII contains the obligation of National Treatment.[10] It states that member countries must treat foreign services

[1] Bradford A. (2023). *Digital Empires: The Global Battle to Regulate Technology*. Oxford University Press.

[2] Li L. (2025). Data Sovereignty and National Security: Governance Challenges and Pathways in the Digital Age. *Global Review of Humanities, Arts, and Society*, *1*, 49.

[3] Pohle J and Thiel T. (2020). Digital Sovereignty. *Internet Policy Review*, *9*.

[4] Diesselhorst M. (1982). Hugo Grotius and the Freedom of the Seas. *Grotiana*, *3*, 11.

[5] Newman H. (2021). *OF PRIVACY and POWER: The Transatlantic Struggle over Freedom and Security.* Princeton University Press.

[6] Pierucci F. (2025). Sovereignty in the Digital Era: Rethinking Territoriality and Governance in Cyberspace. *Digital Society*, *4*.

[7] Dan Jerker B Svantesson. (2017). *Solving the Internet Jurisdiction Puzzle*. Oxford University Press.

[8] General Agreement on Trade in Services (adopted 15 April 1994, entered into force 1 January 1995) 1869 UNTS 183 ('GATS').

[9] ibid art II.

[10] ibid art XVII.

and service providers at least as favorably as their domestic counterparts. However, this is subject to the conditions and limitations specified in their schedules of commitments.

Despite these foundational principles, the GATS framework is structurally ill-equipped to address the realities of the modern digital economy for several reasons.[1] First, it was negotiated and concluded before the widespread commercialisation of the internet, and thus contains no specific provisions on digital commerce, data circulation and data localisation. Second, it is not clear whether digital products ought to be regarded as "merchandise" under the General Agreement on Tariffs and Trade or as "services" under the GATS. This differentiation has substantial legal and economic ramifications. Third, the core GATS duties of market entry and national treatment are applicable only to those service sectors that a member has explicitly included in its schedule of commitments. This approach of using a positive list means that many digitally service sectors remain outside the scope of binding liberalisation commitments for a majority of the WTO members. This legal vacuum has allowed the divergent models of data sovereignty to flourish with few multilateral constraints, setting the stage for direct conflict.

3. Manifestations of Conflict in EU-China Cross-Border Data Flow Policies

3.1 Different Value Orientations Toward Data Sovereignty in Digital Trade

3.1.1 The China Model: Security-Oriented Approach

China's approach to data governance is explicitly state-centric, driven by the dual objectives of national security and economic development. The guiding philosophy is "cyber sovereignty", which treats cyberspace as a domain subject to the same principles of state control as physical territory. Within this framework, data is viewed not primarily as a private asset or personal right, but as a strategic national resource and a "fifth factor of production" in conjunction with land, labour, capital and technology.

The model is operationalized through a complex legal system, comprising the Personal Information Protection Law (PIPL), the Data Security Law (DSL) and the Cybersecurity Law (CSL). The system creates a hierarchical framework for data categorization according to its significance to the national interest. The DSL delineates "important data" and "core national data". The latter pertains to information pertaining to national security, the vital arteries of the national economy, crucial elements of people's daily lives and substantial public interests. Such kind of data is subject to the most stringent regulations, including mandatory data localisation requirements for Critical Information Infrastructure (CII) operators and stringent security assessments carried out by the Cyberspace Administration of China (CAC) before any transfer across national borders. This essentially establishes a data export licensing method, reflecting the core principle that the state has final rights of control over data generated within the border.[2]

3.1.2 The EU Model: Human Rights-Oriented Approach

The EU model operates on a human rights-oriented strategy. It assigns significant importance to the safeguarding of basic human rights. It is committed to ensuring that these rights are respected even outside its geographical limits, either through the domestic law of the recipient country or by specific contractual arrangements.

This principle was tested many times by the transatlantic data flow frameworks. The EU-US Safe Harbor Agreement (invalidated by the Court of Justice of the European Union in 2015),[3] and its successor, the Privacy Shield (struck down in 2020), were both found to be insufficient.[4] These legal rulings were spurred by concerns about the inappropriate utilization of EU citizens' data and the broad access to such data by US intelligence services, all without providing a degree of protection that was "substantially equivalent".

The implementation of the General Data

[1] Irion K, Yakovleva S and Bartl M. (2016). Trade and Privacy: Complicated Bedfellows? How to Achieve Data Protection-Proof Free Trade Agreements. *SSRN Electronic Journal*.

[2] Angela Zhang. (2021). *CHINESE ANTITRUST EXCEPTIONALISM: How the Rise of China Challenges Global Regulation*. Oxford University Press.

[3] Case C-362/14 *Maximillian Schrems v Data Protection Commissioner* [2015] ECLI:EU:C:2015:650, paras 8-10.

[4] Case C-311/18 *Data Protection Commissioner v Facebook Ireland and Maximillian Schrems* (Schrems II) [2020] ECLI:EU:C:2020:559, paras 12-15.

Protection Regulation (GDPR) [1] in 2018 represented a pivotal juncture, codifying the EU's strict data protection legal standards into a directly applicable regulation. Under the GDPR, when personal data is transferred outside the EU, it is required to adhere to specific binding protective measures and data protection stipulations. Crucially, it prohibits data transfers to jurisdictions where public authorities can access such data without meaningful limitations and where individuals lack effective legal recourse. Compared to the 1995 Directive[2], the GDPR significantly improves data protection requirements, marking a milestone in the fields of data security and personal privacy.

3.2 Digital Trade Liberalisation as a Threat to Data Sovereignty

For states that prioritise data sovereignty, unfettered digital trade poses a direct threat to regulatory autonomy and national interests. The borderless digital market promoted by free-trade poses challenges to the state's ability to enforce domestic laws, safeguard citizens' privacy and maintain national security. This perceived loss of control is a primary driver of data sovereignty measures.

The core of this threat lies in the inherent sensitivity of data. Unrestricted cross-border data flows imply that sensitive information, ranging from personal health records to critical infrastructure data, may be transmitted to and stored in foreign jurisdictions, resulting in major security risks. States fear such data could be accessed by foreign intelligence agencies or subjected to inferior legal standards. For developing states, these anxieties are amplified by fears of "digital colonialism", where the economic value of domestic data is extracted by dominant foreign technology firms, exacerbating existing economic disparities. [3] Consequently, many of the developing states view unfettered data liberalisation as a mechanism that disproportionately benefits a few technologically advanced economies at the expense of broader global security and development. Thus, when entities such as China or the EU implement controls on data outflows, these actions are often not regarded as protectionism but as essential exercises of sovereignty necessary to safeguard citizens and strategic interests.

3.3 Data Localisation as a Barrier to Digital Trade

Following the exposure of major data security incidents, such as Edward Snowden's revelations, global concerns over data security have intensified. In response, numerous countries have enacted data localisation policies. Consequently, non-tariff barriers, traditionally seen in the trade of goods, now emerge in digital trade.[4]

These localisation measures negatively impact the liberalisation in several respects. First, they create significant barriers to market entry and increase compliance costs for multinational corporations. Data localisation regulations stipulate that data has to be kept and processed within the territorial boundaries of a nation. To meet this requirement, multinational corporations must either invest heavily in establishing local servers and data centers or outsource these operations to domestic service providers, both of which raise their costs. A prominent example is China's CSL, which stipulates that operators of CII are required to store all personal details and significant data gathered within China on domestic servers. In cases where data needs to be transferred overseas, it has to undergo a rigorous security evaluation by the CAC. This has become the de facto standard across numerous industries, including finance, energy and transportation, rendering cross-border data transfer the exception rather than the norm. From a trade law perspective, this policy constitutes a clear barrier under the GATS. For multinational corporations, such a regulation can be seen as a violation of market access commitments, as it imposes conditions on the cross-border supply of services that were not specified in the agreement. This, in turn, unfairly hinders their entry into the Chinese market and prevents them from competing on par with domestic

[1] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Date Protection Regulation) [2016] OJ L 119/1 ('GDPR').

[2] Directive 95/46/E C of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data [1995] OJ L 281/31.

[3] Couldry N and Mejias UA. (2019). *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism*. Stanford University Press.

[4] Meltzer JP. (2015). The Internet, Cross-Border Data Flows and International Trade. *Asia & the Pacific Policy Studies*, 2, 90.

service suppliers. Second, these measures increase operational costs and reduce trade opportunities for all parties. The free flow of data across borders is instrumental in lowering the costs for businesses to find trading partners and expand overseas. In contrast, data localisation complicates the connection between importers and exporters. [1] This complexity effectively shrinks potential market size and diminishes companies' ability to use data analytics to identify new customers and seize trade opportunities.

4. Pathways to Coordinate Data Sovereignty and Digital Trade Liberalisation

4.1 Philosophical Coordination: Inclusive Development Based on the Values of Global Justice

The construction of a global legal framework is an active and interactive process of coordinated action, joint participation and mutual respect among sovereign states, rather than a hegemonic effort where states compete for dominance and marginalise others. [2] Data security concerns national, public and citizen interests, making international data cooperation a common need for all countries.

First, this requires respecting cultural differences and resisting data hegemony.[3] In the context of globalisation, digital trade has promoted cultural exchange, but it has also triggered conflicts due to differing cultures and value systems across regions. Therefore, international data cooperation must, while respecting and safeguarding the data sovereignty of all nations, construct a global data governance framework that acknowledges and embraces cultural pluralism. All countries should actively promote positive interactions in the digital space and enhance cross-cultural understanding.

Simultaneously, it is imperative to address unequal rights in digital trade. The de facto "data hegemony" restricts the digital economic development of developing countries through extraterritorial measures or market monopolization. To counter these challenges, affected countries should actively coordinate through bilateral or multilateral channels with other countries, regions and international organisations with similar concerns, in order to jointly resist unilateralism and hegemonism.

Second, coordination must involve safeguarding fundamental human rights and strengthening security governance.[4] The ultimate goal of data governance is to promote and protect human rights, and an effective governance model requires comprehensive measures that give equal weight to both technology and law. At the national level, all countries need to strengthen digital infrastructure and develop indigenous technology to narrow the digital divide. Meanwhile, domestic laws must strictly regulate acts that infringe personal privacy and data safeguarding. At the international level, while enjoying the dividends of the digital economy, all countries must jointly assume governance responsibilities, ensuring that the concept of the rule of law pervades every aspect of data-related operations. At the legislative level, a compatible and interoperable system of digital rules should be built through international coordination and cooperation. At the enforcement level, strategic cooperation on cybersecurity should be deepened, establishing effective mechanisms for intelligence sharing and joint operations. At the judicial level, international judicial assistance should be strengthened to jointly combat transnational cybercrime, thereby fostering a stable, transparent and secure global cyber legal environment.

4.2 Rule Coordination: Perfecting the International Legal System for Digital Trade

4.2.1 Formulating Internationally Unified Digital Trade Rules

For a long time, the global digital trade has been dominated by developed states. However, the digital capabilities of developing countries have shown accelerated growth in recent years, accompanied by increasing calls to share in economic and security benefits. This underscores the need for coordination by a unified international organization. The current landscape of international organisations is complex and fragmented across monetary,

[1] AARONSON S. (2015). Why Trade Agreements Are Not Setting Information Free: The Lost History and Reinvigorated Debate over Cross-Border Data Flows, Human Rights, and National Security. *World Trade Review*, *14*, 671.

[2] Kingsbury B, Krisch N and Stewart RB. (2005). The Emergence of Global Administrative Law. *SSRN Electronic Journal*. <https://scholarship.law.duke.edu/cgi/viewcontent.cgi?article=1361&context=lcp> accessed 20 October 2025.

[3] Cohen JE. (2016). The Regulatory State in the Information Age. *Theoretical Inquiries in Law*, *17*.

[4] Dencik L. (2025). "Rescuing" Data Justice? Mobilising the Collective in Responses to Datafication. *Information, Communication & Society*, 1.

investment and trade domains, making it difficult to form a coherent system for data protection. In this context, the WTO is the ideal venue to assume this responsibility. It has extensive experience in rule-making, and it can play a positive role in multilateral negotiations among member states. Furthermore, major economies within the WTO should take the lead. For example, China, Russia, the US and the EU need to break the existing framework and spearhead the drafting of a new multilateral agreement specifically for digital trade. Through negotiation, they can regulate cross-border digital trade, address various legal issues in separate clauses and apply them to specific WTO members.

4.2.2 Establishing Security Exception Clauses for Digital Trades

A more effective approach to rule coordination involves refining the use of exception clauses within digital trade agreements. Both regional and multilateral agreements, like the GATS, the Regional Comprehensive Economic Partnership (RCEP) and the Comprehensive and Progressive Agreement for Trans-Pacific Partnership (CPTPP), include stipulations that enable states to diverge from liberalisation obligations to protect essential public interests.[1] For example, Article XIV of the GATS bis offers a security exemption that permits members to enforce methods deemed "necessary for the protection of essential security interests".[2] Historically, this clause has been largely self-judging, granting significant deference to national governments. However, the WTO jurisprudence, such as the Russia-Measures concerning Traffic in Transit panel report,[3] clarified that the application of security exceptions is subject to a review carried out in good faith and cannot be completely immune from examination.[4]

To prevent abuse and ensure predictability, future digital trade agreements should clarify the scope and application of security exceptions.[5] For example, states could be required to demonstrate a clear and direct link between a data-restrictive measure and a specific, identifiable security threat, rather than relying on vague or broad economic security claims. Agreements could also specify procedural protection, such as notification requirements, transparency obligations and periodic review of security measures. Furthermore, a necessity test could be adopted, requiring states to show that no less restrictive alternative is available to achieve the security objective. Additionally, exceptions could be limited to clearly defined circumstances, such as cyberattacks, threats to critical infrastructure or emergencies in international relations, and be subject to independent dispute settlement review.

4.3 Regulatory Coordination: Strengthening the International Regulatory Mechanism for Digital Trade

4.3.1 Promoting the Integrated Development of Existing Regulatory Models

The slow progress of negotiations for the Trade in Services Agreement and the Transatlantic Trade and Investment Partnership indicates that, given the significant divergences in values and interests among countries, it is extremely challenging to negotiate a universally applicable regime for cross-border data flows.[6] Such attempts are costly and often fail to reach consensus. Therefore, rather than starting from scratch and risking further stalemate, a more pragmatic path is to build upon the existing regulatory models widely accepted by the international community, and to promote interoperability and mutual recognition between different regional paradigms through strengthened cooperation.

Currently, the mainstream global regulations for cross-border data flows follow two typical models. One is the EU model, represented by the GDPR, which relies on "adequacy decisions" for countries or regions supplemented by Standard Contractual Clauses and Binding

[1] Svetlana Yakovleva and Kristina Irion. (2020). Pitching trade against privacy: reconciling EU governance of personal data flows with external trade. *International Data Privacy Law*, *10*, 201.

[2] GATS, art XIV *bis*.

[3] WTO, *Russia: Measures Concerning Traffic in Transit* (26 April 2019) WT/DS512/R.

[4] Lapa V. (2020). The WTO Panel Report in Russia -Traffic in Transit: Cutting the Gordian Knot of the GATT Security Exception? *Questions of International Law, Zoom-in*, *69*, 5. <https://www.qil-qdi.org/wp-content/uploads/2020/05/02_WTO-Security-exceptions_LAPA_FIN.pdf> accessed 20 October 2025.

[5] Wenjia Zuo. (2024). General Exceptions in the Digital Trade Environment: Challenges and Reforms under Article 20 of GATT and Article 14 of GATS. *Journal of Education, Humanities and Social Sciences*, *39*, 77.

[6] Lomotey RK, Kumi S and Deters R. (2022). Data Trusts as a Service: Providing a Platform for Multi-Party Data Sharing. *International Journal of Information Management Data Insights*, *2*, 100075.

Corporate Rules (BCRs).[1] The second model is the APEC, exemplified by the Asia-Pacific Economic Cooperation Cross-Border Privacy Rules (CBPR) system, which is a co-regulatory model based on the accountability of organisations.[2]

Although rooted in different legal traditions and differing in regulatory intensity, mandatory nature and mechanism flexibility, the two models are not fundamentally incompatible.[3] In essence, both aim to bridge regulatory gaps between jurisdictions and establish acceptable baseline protection standards for the secure and free flow of data within their regions. Specifically, there are points of convergence. First, there is an overlap in their scope of application. The BCRs primarily govern internal data transfers within multinational corporations (MNCs), while the CBPR applies to enterprises in the Asia-Pacific region, including MNCs.[4] Second, their core mechanisms are highly similar. Both require organisations to adopt internal privacy policies that comply with their respective standards. Third, their fundamental principles are aligned in spirit. The basic principles of data processing in the GDPR are highly consistent with the nine core principles of the APEC Privacy Framework, such as purpose limitation and data security. Fourth, both frameworks serve as a baseline. They set a "floor" for data protection standards, not a "ceiling", allowing participating jurisdictions and enterprises to adopt stricter measures beyond the baseline.[5]

In fact, the EU's Article 29 Working Party and the APEC Data Privacy Subgroup have long engaged in dialogue on cross-border enforcement cooperation. The two parties established a joint working group to explore the possibility of achieving mutual recognition and compatibility between the two frameworks based on common principles. They have jointly published "Common Referential on the EU System and APEC System Structures". This reference document provides a detailed analysis of the compliance and certification requirements for the BCR and the CBPR, and offers an informal practical checklist for enterprises seeking joint certification.

If the two models eventually achieve institutional integration, data could potentially flow freely between CBPR-certified and BCR-approved organisations, exempting them from duplicate certification. Although this mutual recognition is still in preliminary stages, the close economic and trade ties between the EU and the APEC region will undoubtedly drive further cooperation.

4.3.2 Establishing a Cooperative Organisation for Inter-Regional Data Supervisory Authorities

To fully realise the potential of cross-border data flows for sustainable digital trade, all stakeholders must act in a coordinated, unified and cross-industry manner. Internet platforms naturally transcend geographical limitations, connecting data subjects across different legal jurisdictions and involving the legal systems of multiple countries.[6] When cross-border data flows give rise to legal disputes, issues such as the extraterritorial effect of domestic laws, choice of law and the extraterritorial enforcement of judgments emerge. Crucially, national data supervisory authorities are limited by their sovereign borders and lack the capacity to effectively supervise data once it has left their country. In view of this, it is difficult to ensure that the rights of data subjects are adequately protected and effectively remedied by relying solely on the data supervision and enforcement agencies of a single country.[7] Therefore, international cooperation mechanisms are crucial to the global governance of cross-border data flows.

This requires the establishment of regional cooperative organisations for data supervision to foster dialogue and cooperation among

[1] GDPR, art 45.

[2] Graham Greenleaf. (2019). Global Convergence of Data Privacy Standards: EU GDPR and APEC CBPR Compared. *International Data Privacy Law*, *34*, 85.

[3] Marfia F, Fornara N and Nguyen T-VT. (2017). A Framework for Managing Data Provider and Data Consumer Semantic Obligations for Access Control. *AI Communications*, *30*, 67.

[4] Zrenner J et al. (2019). Usage Control Architecture Options for Data Sovereignty in Business Ecosystems. *Journal of Enterprise Information Management*, *32*, 477.

[5] Zhu J. (2021). The Personal Information Protection Law: China's Version of the GDPR? *Columbia Journal of Transnational Law Bulletin*.

[6] Sullivan C. (2019). EU GDPR or APEC CBPR? A Comparative Analysis of the Approach of the EU and APEC to Cross Border Data Transfers and Protection of Personal Data in the IoT Era. *Computer Law & Security Review*, *35*, 380.

[7] Cappiello C et al. (2019). Data Ecosystems: Sovereign Data Exchange among Organizations (Dagstuhl Seminar 19391). *Dagstuhl reports*, *9*, 134.

regulatory authorities. [1] In recent years, recognising the complexity of regulating cross-border data flows, various jurisdictions including the EU, South Korea, Japan and Singapore have established data supervisory authorities. Against this background, a cooperative organization among data supervisory authorities can be constructed. Through regular meetings and other mechanisms, it could promote in-depth discussions among members on new trends and regulatory policies in the field of cross-border data, facilitate consensus on data processing standards, and provide member parties with an international platform integrating information sharing, enforcement coordination and cooperation.

Given that each authority has independent enforcement powers within its jurisdiction, member parties might consider ceding specific rights to such a cooperative organization, endowing it with limited supranational enforcement capabilities. As a result, the body would have the authority to assess, investigate and even punish specific cross-border data activities, and take the lead in the construction of a complementary dispute-resolution mechanism.

4.3.3 Coordinating Principles for the Extraterritorial Application of Domestic Rules

There is an inherent contradiction between the globality of data flow and the regionality of data legislation, which makes jurisdictional conflicts an inevitable problem. To safeguard the rights and welfare of domestic data subjects and manage data resources, states often extend the extraterritorial reach of their domestic laws. A prime illustration of this is the long-arm jurisdiction of the GDPR.[2] This extraterritorial impact is not limited to the data processing operations of entities based in the EU. It also encompasses the activities of non-EU entities that either provide goods or services to data subjects within the EU or monitor such data subjects.[3] This unilateral assertion of jurisdiction significantly expands administrative enforcement beyond traditional territorial limits, raising legitimacy concerns. First, it may constitute undue interference with the law enforcement sovereignty of other nations, conflicting with the principles of sovereign equality and international comity. Second, this model does not satisfy the practical requirements of the "effects principle". Without judicial assistance and administrative cooperation from the host country, EU supervisory authorities effectively lack the ability to complete investigations and enforcement abroad. If forced implementation occurs, such supervision may become nominal and could provoke trade retaliation or diplomatic disputes. Third, overly broad jurisdictional claims impose high operational and regulatory costs on the EU itself.[4]

Therefore, a system of "jurisdiction by agreement" offers an important solution to this regulatory predicament.[5] To manage conflicts arising from extraterritorial application of national rules, parties could reach consensus on the methods and limits of jurisdiction through international negotiation, taking into account all parties' interests and mutually ceding some powers.

Two specific approaches to coordination are possible. First, the multilateral treaty approach. Countries with significant cross-border data exchanges could jointly sign a "Mutual Recognition and Enforcement of Jurisdiction Agreement". Such an agreement should clearly define the conditions, scope, limits and methods for exercising jurisdiction. It also needs to establish a periodic review mechanism to dynamically adjust the treaty provisions and enforcement standards based on actual cases. Second, the domestic law approach. When drafting domestic data regulations, countries could introduce a mechanism of jurisdictional deference. For example, depending on the nature of a dispute, the circumstances of the parties and the degree of interest and concern, jurisdiction could be ceded to the country with the closest connection, the most convenient jurisdiction or the greatest interest.

5. Conclusion

Through the two different legal frameworks of China and the EU, this study has demonstrated

[1] Arnell P et al. (2021). Police Cooperation and Exchange of Information under the EU–UK Trade and Cooperation Agreement. *New Journal of European Criminal Law*.

[2] GDPR, art 3.

[3] Kuner C. (2021). Territorial Scope and Data Transfer Rules in the GDPR: Realising the EU's Ambition of Borderless Data Protection. *SSRN Electronic Journal*.

[4] Jerker D. (2013). *Extraterritoriality in Data Privacy Law*. Ex Tuto.

[5] Berman PS. (2014). From Legal Pluralism to Global Legal Pluralism. *SSRN Electronic Journal*.

the profound and multidimensional conflict between the principles of data sovereignty and digital trade liberalisation. Both regimes set up barriers to cross-border data flows. The pathways to coordination lie not in global unification, but in pragmatic and multi-level interaction. [1] This involves leveraging any existing flexibility in trade law and learning from innovative balancing mechanisms in new regional agreements. By encompassing legitimate regulatory diversity, the inherent tension can be managed, preserving the immense economic and social benefits of a connected, open and reliable global digital ecosystem.

References

Primary Sources

Case C-311/18 Data Protection Commissioner v Facebook Ireland and Maximillian Schrems (Schrems II) [2020] ECLI:EU:C:2020:559

Case C-362/14 Maximillian Schrems v Data Protection Commissioner [2015] ECLI:EU:C:2015:650

Directive 95/46/E C of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data [1995] OJ L 281/31

General Agreement on Trade in Services (adopted 15 April 1994, entered into force 1 January 1995) 1869 UNTS 183

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Date Protection Regulation) [2016] OJ L 119/1

Secondary Sources

AARONSON S. (2015). Why Trade Agreements Are Not Setting Information Free: The Lost History and Reinvigorated Debate over Cross-Border Data Flows, Human Rights, and National Security. *World Trade Review*, *14*, 671.

Appleton B. (2025). Digital Sovereignty vs. Trade Liberalization: India's Algorithm Disclosure Dilemma. Balsillie Case Studies. <https://balsilliecases.ca/case-study/digital-sovereignty-vs-trade-liberalization-indias-algorithm-disclosure-dilemma/>

Arnell P et al. (2021). Police Cooperation and Exchange of Information under the EU–UK Trade and Cooperation Agreement. *New Journal of European Criminal Law*.

Berman PS. (2014). From Legal Pluralism to Global Legal Pluralism. *SSRN Electronic Journal*.

Bradford A. (2023). *Digital Empires: The Global Battle to Regulate Technology*. Oxford University Press.

Cappiello C et al. (2019). Data Ecosystems: Sovereign Data Exchange among Organizations (Dagstuhl Seminar 19391). *Dagstuhl reports*, *9*, 134.

Cohen JE. (2016). The Regulatory State in the Information Age. *Theoretical Inquiries in Law*, *17*.

Couldry N and Mejias UA. (2019). *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism*. Stanford University Press.

Dencik L. (2025). "Rescuing" Data Justice? Mobilising the Collective in Responses to Datafication. *Information, Communication & Society*, 1.

Diesselhorst M. (1982). Hugo Grotius and the Freedom of the Seas. *Grotiana*, *3*, 11.

Gao HS. (2021). Data Sovereignty and Trade Agreements: Three Digital Kingdoms. *SSRN Electronic Journal*.

Greenleaf G. (2018). Global Convergence of Data Privacy Standards and Laws: Speaking Notes for the European Commission Events on the Launch of the General Data Protection Regulation (GDPR) in Brussels & New Delhi, 25 May 2018. *SSRN Electronic Journal*.

Irion K and Yakovleva S. (2020). Pitching trade against privacy: reconciling EU governance of personal data flows with external trade. *International Data Privacy Law*, *10*, 201.

Irion K, Yakovleva S and Bartl M. (2016). Trade and Privacy: Complicated Bedfellows? How to Achieve Data Protection-Proof Free Trade Agreements. *SSRN Electronic Journal*.

[1] Neal Kushwaha, Przemyslaw Roguski and Bruce W Watson, 'Up in the Air: Ensuring Government Data Sovereignty in the Cloud' [2020] 2020 12th International Conference on Cyber Conflict.

Jerker D. (2013). *Extraterritoriality in Data Privacy Law*. Ex Tuto.

Kingsbury B, Krisch N and Stewart RB. (2005). The Emergence of Global Administrative Law. *SSRN Electronic Journal.* <https://scholarship.law.duke.edu/cgi/viewcontent.cgi?article=1361&context=lcp>

Kuner C. (2021). Territorial Scope and Data Transfer Rules in the GDPR: Realising the EU's Ambition of Borderless Data Protection. *SSRN Electronic Journal*.

Kushwaha N, Roguski P and Watson BW. (2020). Up in the Air: Ensuring Government Data Sovereignty in the Cloud. 2020 12th International Conference on Cyber Conflict.

Lapa V. (2020). The WTO Panel Report in Russia -Traffic in Transit: Cutting the Gordian Knot of the GATT Security Exception? *Questions of International Law, Zoom-in*, *69*, 5. <https://www.qil-qdi.org/wp-content/uploads/2020/05/02_WTO-Security-exceptions_LAPA_FIN.pdf>

Lateef MA. (2025). Digital Sovereignty in Global Trade: Analysing WTO Governance of Data Flows. *Beijing Law Review*, *16*, 875.

Li L. (2025). Data Sovereignty and National Security: Governance Challenges and Pathways in the Digital Age. *Global Review of Humanities, Arts, and Society*, *1*, 49.

Lomotey RK, Kumi S and Deters R. (2022). Data Trusts as a Service: Providing a Platform for Multi-Party Data Sharing. *International Journal of Information Management Data Insights*, *2*, 100075.

Marfia F, Fornara N and Nguyen T-VT. (2017). A Framework for Managing Data Provider and Data Consumer Semantic Obligations for Access Control. *AI Communications*, *30*, 67.

Meltzer JP. (2015). The Internet, Cross-Border Data Flows and International Trade. *Asia & the Pacific Policy Studies*, *2*, 90.

Newman H. (2021). *OF PRIVACY and POWER: The Transatlantic Struggle over Freedom and Security.* Princeton University Press.

Pierucci F. (2025). Sovereignty in the Digital Era: Rethinking Territoriality and Governance in Cyberspace. *Digital Society*, *4*.

Pohle J and Thiel T. (2020). Digital Sovereignty. *Internet Policy Review*, *9*.

Sullivan C. (2019). EU GDPR or APEC CBPR? A Comparative Analysis of the Approach of the EU and APEC to Cross Border Data Transfers and Protection of Personal Data in the IoT Era. *Computer Law & Security Review*, *35*, 380.

Svantesson DJB. (2017). *Solving the Internet Jurisdiction Puzzle*. Oxford University Press.

Zhang A. (2021). *CHINESE ANTITRUST EXCEPTIONALISM: How the Rise of China Challenges Global Regulation*. Oxford University Press.

Zhu J. (2021). The Personal Information Protection Law: China's Version of the GDPR? *Columbia Journal of Transnational Law Bulletin*.

Zrenner J et al. (2019). Usage Control Architecture Options for Data Sovereignty in Business Ecosystems. *Journal of Enterprise Information Management*, *32*, 477.

Zuo W. (2024). General Exceptions in the Digital Trade Environment: Challenges and Reforms under Article 20 of GATT and Article 14 of GATS. *Journal of Education, Humanities and Social Sciences*, *39*, 77.

PIONEER
ACADEMIC PUBLISHING LIMITED

Studies in Law and Justice
ISSN 2958-0382
www.pioneerpublisher.com/slj
Volume 4 Number 5 October 2025

Effects of Biometrics on Public Security in Brazil: An Analysis of Facial Recognition and Algorithmic Racism

Pablo Ornelas Rosa[1], Clécio José Morandi de Assis Lemos[2], Aknaton Toczek Souza[3] & Eudson Ferreira Bento[4]

[1] Doctor of Social Sciences; Professor, Universidade Vila Velha (UVV) & University Center of Vale Cricaré (UNIVC); Productivity Grant Holder, Fundação de Amparo à Pesquisa do Espírito Santo (FAPES), Brazil
[2] Doctor in Law; Professor, Federal University of Espírito Santo (UFES) & University Center of Espírito Santo (UNESC), Brazil
[3] Doctor of Sociology and Law; Professor, Catholic University of Pelotas (UCPEL), Brazil
[4] Master in Public Security; Chief of the Civil Police of Espírito Santo (ES), Brazil
Correspondence: Pablo Ornelas Rosa, Doctor of Social Sciences; Professor, Universidade Vila Velha (UVV) & University Center of Vale Cricaré (UNIVC); Productivity Grant Holder, Fundação de Amparo à Pesquisa do Espírito Santo (FAPES), Brazil.

doi:10.56397/SLJ.2025.10.08

Abstract

This article presents a literature review on the relationship between racism and the use of facial recognition technology by Brazilian Public Security. The text is organized into four sections: the first presents phenomena such as platformization, the datafication of life, data colonialism, platform capitalism, and surveillance; the second sets out some characteristics of this technology; the third highlights studies on racial discrimination through algorithms; finally, the fourth presents the main consequences of algorithmic racism in Public Security. It concludes that there is a need for a broad debate on these discriminatory algorithmic practices, in order to avoid the violation of fundamental rights and guarantees.

Keywords: racism, algorithm, public security, fundamental rights

1. Introduction

Tools such as artificial intelligence, digital certification, the Internet of Things, among many others, are just some examples of how technology has been transforming the everyday lives of individuals, companies, and the State itself. As it becomes embedded in our practices, producing changes in behavior, technology also presents impasses. One of these questions concerns the new guise that racism takes on, including involuntarily, in this ocean of bytes, contaminating it with its structural characteristics, particularly algorithmic blocking operations targeting Black people and the use of

digital robots (bots)[1] for supremacist discourses, most of the time operating without being noticed (SILVA, 2020).

The subtlety of this new form of discrimination arises from the fact that we have come to live in a society characterized by the intensive use of software that becomes the primary form of social interaction, which implies a broad mobilization of algorithms used for predictive purposes (SILVEIRA, 2016; DA EMPOLI, 2019; ZUBOFF, 2020). The normalized use of mobile phones, tablets, and other gadgets indicates the growing presence of these technologies in our everyday lives, intensely permeating our relationships.

However, neither software nor the algorithms contained in it operate in a neutral manner. On the contrary, they produce effects because they are created and developed by human beings with certain purposes in mind. Thus, although they are immaterial and apparently invisible, they have a starting point and a goal that may express discrimination, even if in an unintentional way.

These facts become even more concerning with the application of facial recognition by Public Security agencies, which makes it essential to study the precedents involving artificial intelligence and racism, including their concepts, history, and structure. Accordingly, this article aims to analyze this new facet of racial discrimination, contributing to its understanding and investigation insofar as it seeks to grasp how it occurs today. For some time now, this method of facial recognition has been surrounded by controversy, especially when studies indicate that such technology is prone to errors in the analysis of the faces of Black people or other minorities, which deserves our concern (SILVA, 2020; BEIGUELMAN, 2021; AMARAL, MARTINS & ELESBÃO, 2021; NOBLE, 2021).

The text that follows is organized as follows: the first section sets out some of the assumptions that address the phenomena of platformization and the datafication of life; the second addresses some of the main characteristics of facial recognition technologies; the third highlights the elements that constitute them; and, finally, the last section presents some of the main consequences of this form of social control, with an emphasis on the field of Public Security.

2. Platform and Surveillance Capitalism

On the way into the twenty-first century, significant transformations took place in the ways in which people began to relate to one another, as well as in how they are constituted as subjects, due to the gradual intensification of the use of digital platforms in their communication, altering access to information, which had previously been marked by the primacy of face-to-face interaction (ROSA, AMARAL & NEMER, 2021, p. 02).

Therefore, it is important to describe the process of platformization and the datafication of life, from whose influence Public Security is not immune. With regard to platformization, it is necessary to highlight the contributions of Poell, Nieborg, and Dijck (2020, p. 05), who define it as a form of "penetration of platform infrastructures, economic processes, and governmental structures into different economic sectors and spheres of life."

Following the same authors, it is possible to clarify that platforms are programmed digital models that act upon interactions between people and complementors, doing so through the systematized collection of data, the use of algorithms, and monetization.

Nevertheless, it is also necessary to highlight other significant contributions on this phenomenon, characterized by the conditioning of human relations on social networks, described sometimes as "surveillance capitalism" (ZUBOFF, 2020) and sometimes as "platform capitalism" (SRNICEK, 2018).

Understanding that we have entered an era characterized by what he called platform capitalism, Nick Srnicek (2018, pp. 44–45) identified five specific types of digital platforms that operate on the basis of distinct business models: (a) advertising platforms, which extract and use user data as products sold to advertisers, as in the cases of Google and Facebook; (b) cloud platforms, owners of business hardware and software that are dependent on the digital and that generate profits according to the needs of their client firms, on the basis of an enormous logistical network, such as Amazon and Web Services; (c) industrial platforms, such as General Eletric and Simens, which produce the hardware and software necessary for the transformation of

[1] "Bots are autonomous applications that run on the Internet while performing some type of predetermined task" (GARRET, 2022).

traditional manufacturing into processes connected to the internet; (d) product platforms, such as Rolls Royce and Spotify, which transform a traditional good into a service and charge a rental fee or subscription fee; and (e) lean platforms, such as Uber and Airbnb, which operate through subcontracting, charging a high cost for their use. The author further clarifies that these categories may coexist within the same company.

According to Zuboff (2020, p. 247), the impact of what she called surveillance capitalism, characterized by the intensive use of digital platforms, is felt in the infrastructures of markets, governance, and, notably, data.

In this last case, it is possible to see that it gives rise to the so-called datafication of life, understood by André Lemos (2021, p. 02) as "forms of transforming actions into quantifiable data, allowing extensive tracking and predictive analyses," with the potential to expand into many other fields, such as politics, the economy, culture, etc., reaching the field of Public Security through its consequent biometric use via facial recognition technologies.

As we deepen our understanding of the datafication of life, it is possible to understand it as a new way of producing knowledge, involving a digital requisition or even translation of the world that makes it possible to exercise a certain control over objects and/or actions, with the aim of simulating and testing them in advanced computer systems operated by artificial intelligence (AI). Thus, we have a new hegemonic way of knowing and managing life on the planet (LEMOS, 2021, p. 197).

In this process, the datafication of life has influenced various forms of knowledge, including scientific knowledge, since it has become evident that data do not function in a neutral way, insofar as they produce biases, favoring a technocratic power operated under the tutelage of specialists in algorithms and with public interests.

Lemos (2021, p. 198) further adds that the datafication of knowledge could promote a power led by an "epistocracy," operated through an "algocracy" grounded in the technical neutrality of algorithmic performativity, which would decide about doing and knowing, insofar as it would introduce into human interactions a kind of lens that, just as mathematics was instrumentalized by Newton in the seventeenth century, could be treated as "the great book of nature."

The understanding of this phenomenon perhaps becomes clearer in the analyses carried out by Siva Vaidhyanathan (2011, p. 40) when he deals with the Googlization of everything. According to the author, "Google collects the gigabytes of personal information and creative content that millions of its users provide free of charge to the network every day, and sells this information to advertisers of millions of products and services." In this way, by noting that Google asserts itself by persuading us that it knows exactly what to do to improve our lives, Vaidhyanathan (2011, p. 29) found that this company has come to determine our behavior, controlling the network without raising any suspicion that it exercises authoritarian practices.

The datafication of life is understood by André Lemos (2021, pp. 199–200) as a new era of digital culture, anchored in some dimensions that can be systematized in the form of: (a) knowledge, as it involves a new production through the extraction and management of data; (b) sociability, since it makes the surveillance and collection of personal information routine; and (c) nature, insofar as it negatively impacts the environment through the way natural goods (especially minerals) are consumed and electronic waste is discarded, in addition to the high energy consumption of data centers. Thus, although these impacts seem to be completely unknown to the public, it is important to remember that:

Data are not found in nature, as Couldry and Mejias (2019) have warned. This is a crucial point of the phenomenon of datafication. They are designed and depend on extraction and storage algorithms. As Tarleton Gillespie (2014) aptly pointed out, data are presented as objective and the algorithms that process them are portrayed as above suspicion and incapable of adopting ideological positions, thus becoming a powerful weapon for overcoming controversies. [...]. However, debates and research are advancing that consider not only the biases and prejudices embedded in data structures but also in the codes and algorithms that carry the outlook of their developers and funders (SILVEIRA, 2016, pp. 159–160).

Once created, data can be extracted within a process known as data colonialism, in a way that

is not very transparent to their owners, such that people's habits become a commercializable resource, as they are essential in the relationship between corporations and platforms and can also be used for political competition (SILVEIRA, 2016), as well as for biometric use through facial recognition technologies. Thus, in this combination of state–corporate actors, coloniality ends up being updated, with new instruments but still perpetuating the same destructive and dehumanizing designs inherent to capitalism (GERVASONI & DIAS, 2023, p. 155).

In this sense, it is possible to consider how the massive use of data has ended up enabling a government of conduct, managed in a complex and targeted way by platforms and their algorithms, in a progressive accumulation of information that will be useful to secure positions of strategic advantage.

Silveira (2016) found that digital platforms increasingly began to create datafication projects aimed at converting any digitizable element into a process of capital reproduction. According to the author, this happens because the relations between producers and consumers of a given product, or even between providers and users of certain services, are gradually instrumentalized by platforms managed through algorithms that allow these relations to be consolidated ever more quickly and in line with advertising interests: "Simultaneously, these algorithmic managers extract data from markets and store them with the aim of expanding the knowledge and control of their platforms" (SILVEIRA, 2016, p. 168).

Although commonly associated with Information and Communication Technology (ICT), the concept of the algorithm dates back to the beginnings of mathematics and exists independently of today's digitalization. Since the time of Egyptian civilization, algorithms were used to create formulas that solved everyday challenges, such as predicting the floods of the Nile River, representing a specific sequence of written steps to solve a particular problem. Today, they remain an essential element in the entire computing process, aimed at mediating human activities and reducing the number of repetitive procedures (ROCHA, PORTO & ABAURRE, 2020).

Algorithms play a fundamental role in the operation of artificial intelligences, being essential for the execution of tasks. Although there is no universally accepted concept of Artificial Intelligence (AI), it is commonly understood as the capacity of machines to reproduce behaviors typical of human beings, grounded in the manipulation of algorithms. Currently, AI is applied in three main areas: machine learning, deep learning, and natural language processing (BON, SCHONS & LOPES-FLOIS, 2023, p. 227).

According to Costa (2021), the use of machine learning programs and their more advanced variant, known as deep learning, has given machines a remarkable ability to evolve through experience, as well as to make decisions autonomously. This means that, after the development of the algorithm, many subsequent steps can be carried out without the need for human intervention.

With regard to facial recognition, understood as the capacity to identify individuals by means of characteristics determined by their faces, several authors adopt an optimistic approach to its use for purposes of social control, arguing that the identification of people through the use of such techniques may become a safe and minimally invasive alternative, as recognized by Pablo Nunes et al. (2016).

In this case, the argument in favor of the use of this type of strategy in the field of Public Security, in general, presupposes the development of technologies driven by facial recognition in association with existing video surveillance systems, "which could operate as effective tools in combating crime, especially in locating and identifying fugitives, criminals, missing persons, etc." (Nunes et al., 2016, p. 114).

However, in promising to fight national crime with a supposedly efficient and objective technological resource, there is a serious risk of adopting it without the necessary critical analysis, disregarding those risks that disproportionately affect certain social groups.

3. Facial Recognition Technology

Facial recognition "is a biometric identification technique, like fingerprinting, in which software maps facial lines and, by means of algorithms, compares them to a digital image, recognizing (or denying) the person's identity" (MAGNO & BEZERRA, 2020, p. 46). Its concept was first developed in the 1960s, when Woodrow Wilson Bledsoe, Helen Chan Wolf, and Charles Bisson

created the first semi-automatic recognition system (TRASLAVIÑA, 2015, p. 55).

Over the course of the 1970s, 1980s, and 1990s, other techniques were added and improved. However, only in 2001, during a Super Bowl game of the National Football League (NFL), were images of fans' faces captured by means of surveillance cameras for later comparison with a database, demonstrating the potential of this technology (NUNES et al., 2016, p. 117).

It is not by chance that, in 2019, in Hong Kong, an autonomous territory of China, participants in protests against that country's government destroyed video surveillance cameras in public areas. This attitude should not be treated as mere vandalism, but as a form of defense against future individual repression, by avoiding being recognized (ELESBÃO, SANTOS & MEDINA, 2020, p. 247).

As for its functioning, facial verification is carried out basically in two stages: the moment of detecting the face itself and the moment of its verification, using, simultaneously or separately, two approaches: the global approach, in which an image of thousands of pixels is reduced to a set of numbers, known as Holistic Methods; and the local approach, in which the "local" characteristics of the face are extracted, such as eyes, mouth, and eyebrows, using their positions on the face, known as Structural or Local Methods (NUNES et al., 2016, pp. 119–120).

According to research presented in the Aguará Project (Otegui et al., 2006, p. 80), the algorithm must take into account aspects that complicate the recognition process, such as: "the person's emotional state, due to the recognition of expressions (sad, happy, angry, etc.); location of relevant features found in the eyes, mouth, eyebrows, chin, ears, etc.; face size; presence of glasses, beard, caps, etc.; facial expression; lighting problems; image conditions; unknown number of faces in the image, etc."

That said, we can affirm that this technology has gradually developed over recent decades, moving toward an increasingly broad and complex mode of operation as it assimilates new variables. This is because the possibility of collecting more data and processing them more quickly has allowed significant advances in the accessibility of such mechanisms, making this device increasingly common for purposes of social control, both in the private and in the public sector.

According to Nunes (2019), Brazil officially adopted the use of facial recognition technologies in the area of Public Security only in 2019, after a year of experimentation in some states of the country, worsening mass incarceration mainly as a result of the arrest of young Black people from Brazilian peripheries. In that year, the state of Bahia was the first to adopt this type of technology during Carnival, resulting in the arrest of 74 people.

Although the promises associated with these biometric technologies are tempting, seeing in the use of facial recognition a way to increase the efficiency of police work, great caution is required in a country where the police are questioned for their racist bias. There is a constant risk that the dangers of racial prejudice in these technologies will be minimized, insofar as it is assumed that the algorithm is "neutral" in the task of selecting potential suspects (NUNES, 2019).

It must be explained that the parts of the body most used in biometrics, whether fingerprints or the face itself, will never be fully analyzed, since only some of their points are selected in order to calculate the probability that they are features of the person registered in the database. If similarity levels below the established 90% are set, this may lead to a large number of identifications, generating a significant quantity of false positives. Conversely, "if the level of similarity required by the algorithm is 99.9%, for example, the likelihood that the system will issue alerts will be very low" (NUNES, 2019, p. 68). It is not difficult to conclude that such false positives would inevitably translate into public humiliation, arbitrary arrests, and clear violations of fundamental rights and guarantees.

The Rede de Observatórios da Segurança has monitored cases of arrests and police stops resulting from the use of facial recognition, as well as projects to implement this form of surveillance and control in the country. According to its reports, it was found that, from March to October 2019, cases of arrests resulting from the use of facial recognition technology were monitored in four Brazilian states: Paraíba, Bahia, Rio de Janeiro, and Santa Catarina. "Of the cases monitored by the Rede, Bahia accounted for 51.7% of arrests, followed by Rio de Janeiro with 37.1%, Santa Catarina with 7.3%, and Paraíba with 3.3%" (NUNES, 2019, p. 69).

Although in some monitored cases, it was difficult to find precise information about the profile of the people arrested or stopped by the police, taken as a whole, that is, for all 66 identified cases, there was information on sex, age, race/color, and motivation. Among those investigated, it was possible to verify that 87.9% of the suspects were men and 12.1% women; the average age of the group under scrutiny was 35 years; and 90.5% of the people were Black and 9.5% were white. With regard to motivation, the highest numbers were for the crimes of drug trafficking and robbery, each with 24.1% (NUNES, 2019, p. 69).

In this case, it seems necessary to emphasize that, while countries such as Belgium have begun to adopt a ban on facial recognition technology, as Nunes (2019) highlights, in Brazil this approach appears to be moving in the opposite direction, insofar as the number of enthusiasts is increasing. States such as Minas Gerais, Espírito Santo, Pará, and the Federal District have already declared that they are in the process of contracting or implementing this type of technology in the field of Public Security. The same seems to be occurring in all the states of the Northeast, driven by projects of Chinese companies that are being implemented in this region.

The federal government has contributed significantly to the expansion of this type of technology, as can be seen in Ordinance No. 793 of 24 October 2019, which regulates the use of money from the National Public Security Fund for the "promotion of the implementation of video surveillance systems with facial recognition solutions, Optical Character Recognition – OCR, the use of artificial intelligence, or others" (NUNES, 2019, p. 69).

Thus, it becomes troubling to consider that, in a country where the basic principles of data transparency in the field of public security have historically been disrespected, and where current projects fully disregard the Lei Geral de Proteção de Dados Pessoais (LGPD – General Data Protection Law), there seems to be no concern with developing accountability mechanisms for facial recognition technologies, nor protocols aimed at ensuring the security of the data collected.

This concern grows when we see that projects involving the use of facial recognition by police forces in some Brazilian states operate in line with the creation of the National Multibiometric and Fingerprint Database proposed by the then Minister of Justice, Sérgio Moro. This database was presented as an important and necessary form of modernization of police practice; however, according to specialists, it has been regarded as a step backwards in terms of efficiency, transparency, and the protection of the population's personal data (NUNES, 2019).

4. Algorithmic Racial Discrimination

Silva argues that algorithms and artificial intelligence, increasingly present in our daily lives through the use of biometrics to unlock smartphones and facial recognition to access certain spaces, can raise various concerns related to prejudice associated with race, gender, social class, location, and neurodivergence. According to the author, such technologies do not operate in a neutral manner, since they entail a process of racialization and algorithmic oppression that results in discriminatory experiences. Thus, programming can be responsible for perpetuating various prejudices and errors (SILVA, 2020).

Although they were conceived with the aim of impartiality, seeking to overcome the limits of human rationality, algorithms absorb the choices, inclinations, and prejudices of their programmers, even if unintentionally, which justifies concern with racial algorithmic discrimination (FRAZÃO, 2021).

In analyzing growing racial discrimination on the World Wide Web, Cardozo (2022) found that Black women are commonly victims of hate speech on social media. According to the author, racial algorithmic discrimination has emerged contemporaneously as the main challenge in confronting this issue, which is consolidated in the infrastructure and interface of digital technologies, in image-processing resources, in content recommendation, among other aspects that highlight the need to discuss the "whiteness" expressed on the internet.

Scholars point to a significant example of algorithmic discrimination in the operation of COMPAS (Correctional Offender Management Profiling for Alternative Sanctions), an artificial intelligence system used by U.S. courts to estimate a defendant's likelihood of recidivism. The criteria assessed—such as place of residence, history of involvement with drugs, family background, and school performance—resulted in a classification of

"high risk" of recidivism for Black people significantly more frequently and in greater numbers than for white individuals. This scenario exposes the prejudices embedded in the algorithms, stemming from the parameters defined by programmers (SOARES et al., 2022).

According to Taute (2020), an algorithm is like a recipe, an instruction that the machine follows and, in order to execute it, it must query a database. If this database contains racial prejudices, many people will be included in and excluded from the process, accentuating disparities. In this regard, it is worth recalling the research carried out by Joy Buolamwini, a 30-year-old Black woman and researcher at the Media Lab of the Massachusetts Institute of Technology (MIT).

When developing a prototype of a smart mirror capable of recognizing the face of the person in front of it and projecting features of inspiring figures such as Serena Williams, Joy Buolamwini attached a camera to capture the image of her face and transmit it to her computer which, by means of a facial recognition algorithm, would identify the person and link them to personalized information. However, when she began the experiment, the prototype did not detect her face, only succeeding after she used a white mask, showing that the color of her skin prevented the system from working (NUNES, 2021).

Another example that should be mentioned concerns the choice of who would succeed actor Daniel Craig in the role of James Bond. At first, it was reported that the selection for this role would be made using Artificial Intelligence and that it would point to a Black woman. However, the final selection contradicted what had been reported in 2019, resulting in the hiring of actor Henry Cavill for the role, another white man. What happened was that the AI had been trained on data from the film industry produced in the Global North, where Black people are minimally represented among protagonists (BEIGUELMAN, 2021).

A study by Tarcízio Silva et al. (2020, p. 30) on labeling failures in Google Cloud Vision, focusing on images of Black women, "showed that the photos recurrently received the label 'wig' whenever their hair was prominent," revealing that the database lacked labels for curly or non-straight hair— a culturally rooted limitation on the part of those responsible for the algorithms. Thus, according to the authors, "this universe of social relations at the base of AIs shows that the supposed misogyny and racism of the algorithms have unmistakable human and political dimensions" (SILVA et al., 2020, p. 33).

As indicated by the National Institute of Standards and Technology (NIST) of the United States government, the algorithms normally used are much less accurate in the facial identification of African American and Asian individuals than in that of white people. In this context, Black women had higher probabilities of being misidentified, thereby perpetuating racist practices under a technological guise (NUNES, 2019).

The work of Christina Baker (2005) also stands out for recognizing that media stereotypes attributed to white and Black women differ substantially, insofar as the images most commonly associated with Black women do not reflect the same affability and submissiveness as those associated with white women, who are not recurrently portrayed, from an imagetic standpoint, as sexually aggressive and animalistic, threatening men in their masculinity.

Amaral, Martins, and Elesbão (2021) mention research carried out on the content of image banks regarding racial patterns of families on digital platforms, which predominantly maintain a profile of white people.

In the Getty Images database, of the 300 images returned for the term "family," 107 were of entirely white families, 24 of entirely Black families, and 22 of interracial families and other races/ethnicities. In the Shutterstock image bank, of the 319 images returned, 214 were of entirely white families, 20 of entirely Black families, and 20 of interracial families and other races/ethnicities. Finally, in the Stock Photos image bank, of the 301 images returned for the term "family," 213 were of entirely white families, 14 of entirely Black families, and 15 of interracial families and other races/ethnicities (AMARAL, MARTINS & ELESBÃO, 2021, p. 07).

Thus, discussion of this topic is urgent, since racial inequalities are potentially reflected through algorithms as an extension of the programmer's opinions, values, and social standards, as exemplified by the way images are made available to users on the internet (AMARAL, MARTINS & ELESBÃO, 2021, p. 05).

5. Facial Recognition and Racism in Public Security

With regard to the application of facial recognition in the area of Public Security, its history goes back to the terrorist attack that took place in the United States on 11 September 2001. It was from this event that the use of this technology for crime prevention in various countries was driven forward, becoming a true milestone (NUNES et al., 2016, pp. 123–124).

Since then, facial recognition has increasingly been treated as a promising technology in the field of Public Security. Through advanced algorithms, it is said to be capable of identifying individuals on the basis of unique characteristics of their faces, comparing them with databases of previously registered images. Thus, according to its enthusiasts, this would make it possible to quickly identify suspects, people wanted by the courts, and individuals involved in criminal activities (MELO & SERRA, 2022).

According to Rola (2022), the most impactful biometric technology today is facial recognition. Unlike other biometric forms, such as fingerprints, iris or retina scans, and voice, facial recognition is fast and discreet in terms of data collection, since it generally does not require the cooperation of the person being identified. In contrast to other biometric modalities that require the individual's consent, facial recognition emerges as an investigative tool.

However, as already noted, facial recognition can be influenced by environmental factors such as lighting, angle of capture, facial expression, pose, makeup, and accessories like glasses and hats. Thus, possible mistakes in facial recognition highlight a very dangerous weakness in the realm of security. In other words, these errors underscore the importance of improving facial recognition algorithms, should they in fact be implemented, through advances in artificial intelligence so as to make them more robust and accurate, meeting police demands without resulting in reckless criminalization.

Accordingly, continuous investment in research and development can help reduce error rates and increase the reliability of facial recognition as a security tool. Moreover, it is essential to ensure that ethics and data protection are taken into account in the implementation of these technologies, seeking a balance between security and individuals' privacy (ROLA, 2022).

According to Francisco, Hurel, and Rielli (2020, p. 17), in light of the procedural flaws mentioned above, many oppose the use of facial recognition technology by Public Security agencies, because scientific research has shown high error margins when analyzing the faces of women and Black people. Despite the slow development of regulatory control, there has been an increase in its incidence in Brazil, given that the use of facial recognition by police forces, municipal guards, and other Public Security bodies has occurred in at least 30 cities across 16 states of the country up to 2022.

Monitoring carried out by Intervozes revealed that, among the 26 mayors of state capitals sworn in in January 2021, 17 presented proposals concerning the use of Information and Communication Technologies in the field of Public Security, including the implementation of facial recognition technology (GOMES & MOURA, 2022).

In turn, several relevant problems in its use have already been reported, such as the one that occurred during a period of testing of facial recognition technology on Copacabana beach. On the second day of the experiment, a woman was recognized as being Maria Lêda Félix da Silva, convicted of homicide and wanted by the police, for which reason she was arrested and taken to the police station. After all the embarrassment inherent in this type of procedure, the woman was released when her family members brought her documents proving that she was not the person flagged by the algorithm.

The case illustrates yet another example in a series of errors produced by these technologies, but with an aggravating factor: Maria Lêda, the "wanted woman," had already been serving her sentence in a prison for four years. In this case, not only did the algorithms fail, but so did the police, who used an outdated database (NUNES, 2021).

The issue takes on particular significance when there is a widespread view that technology, along with science, is objective, which makes it harder to understand. This supposed objectivity is contestable, since those who fund and manage these systems play an important role in the outcome. This is an area undergoing rapid growth, without an adequate overall political and ethical debate, thus producing what we may call algorithmic racism, understood as "the way

in which the current arrangement of technologies and sociotechnical imaginaries in a world shaped by white supremacy reinforces the racialized ordering of knowledge, resources, space, and violence to the detriment of non-white groups" (SILVA, 2020; SILVEIRA, 2022).

It is therefore worth recalling the case of the city of Oakland, in the U.S. state of California, whose City Council prohibited, in 2019, the use of facial recognition by public agencies, including the police itself, due to the risks it poses to city residents, with the possibility of misidentifying individuals and the subsequent misuse of force, wrongful arrests, and persecution of minorities (MAGNO & BEZERRA, 2020, p. 51).

Recognizing that the risks and harms associated with the use of facial recognition technology outweigh its possible benefits, the city of San Francisco became the first U.S. municipality to ban its use by Public Security agents, in May 2019. According to the arguments presented by legislators, facial recognition allows for the exacerbation of social injustice and threatens to heighten existing risks.

In this context, advocates of banning the use of this type of technology by Public Security forces point out that the algorithmic models used to train facial recognition technology are developed mostly by white men, which significantly increases the likelihood of misidentifying Black people. Furthermore, in order to train this type of technology, the system must scan the faces of those who circulate in public spaces, even if these people are unaware of it, expanding a state of constant surveillance (GOMES & MOURA, 2022).

Thus, although it deals with highly complex concerns that have sparked ethical and political debates, this is still a developing area without the necessary critical approach. The risks of impacting fundamental rights guaranteed by the Constitution of the Brazilian Republic are considerable, especially if we consider that injustices in the field of public security directly entail vexatious public exposure of one's image, restrictions on freedom, and, eventually, even death.

6. Final Considerations

This article has presented a literature review on the use of facial recognition technology in the field of Public Security, associating it with the perpetuation of discriminatory practices through so-called algorithmic racism. With this aim, it sought to provide basic notions of platformization, the datafication of life, data colonialism, surveillance and platform capitalism, algorithmic racial discrimination, and so on, as well as to raise questions about its use by Brazilian Public Security agencies.

In this regard, it was possible to observe that algorithms are not impartial by nature and can, in fact, incorporate the biases of their creators or of the data sets used during their training. At this stage, the performance of the algorithm may present a biased tendency, since the prejudices present in the training data will be reflected in its decisions and actions.

This issue is particularly important when it comes to applications that directly affect people's lives, especially through the use of decision-making systems based on this type of biometric technology. If the data used to train these algorithms contain prejudices—whether of gender, race, social class, or any other kind—it is likely that the system will reproduce and even amplify these patterns in its decisions.

In this way, power is exercised subtly: the capacity to kill or to let live is exercised without being noticed, through a technology that does not operate by neutral use of its data. It is therefore necessary to understand and limit its application, under penalty of subjecting part of society to a new tool of racial discrimination, with a broader dissemination of oppressive practices.

Understanding the balance between the right to public security and the right to due process of law is absolutely necessary, given the imperative of guaranteeing the constitutional right not to be subjected to unjustified unequal treatment. In a context in which the State provides the public service of protecting the collectivity, utmost caution is required in light of the history of countless acts of violence and racial discrimination in a country marked by a slaveholding legacy.

References

AMARAL, Augusto Jobim do; MARTINS, Fernanda; ELESBÃO, Ana Clara. (2021, out./dez). Racismo algorítmico: uma análise da branquitude nos bancos de imagens digitais. *Revista Pensar, Fortaleza*, *26*(4), p. 1-9, Disponível em: https://ojs.unifor.br/rpen/article/view/11806 Acesso em: 23 jul. 2024.

BAKER, Christina N. (2005, jan.). Images of women's sexuality in advertisements: a content analysis of black and white oriented women's and men's magazines. Sex Roles, 52(1/2), p. 13-27, Disponível em: https://link.springer.com/article/10.1007/s11199-005-1190-y Acesso em: 24 jul. 2024.

BARROS, Isabel Maria Pereira Paes de; SILVA, Isabel Inês Bernardino de Souza. (2020, jul.). Utilização do reconhecimento facial eletrônico por empresas para identificação de suspeitos: segurança ou violação do estado democrático de direito? *Revista Transgressões: Ciências Criminais em debate*, *8*(1), Disponível em: https://periodicos.ufrn.br/transgressoes/article/view/19909 Acesso em: 23 jul. 2024.

BEIGUELMAN, Giselle. (2021). *Políticas da imagem: vigilância e resistência na dadosfera*. São Paulo: Ubu Editora.

CARDOZO, Glenda Dantas. (2022, dez.). A atuação estratégica de mulheres negras no combate às brechas digitais de gênero e raça. *Internet & Sociedade*, *3*(2), p. 5-19.

COSTA, Diego Carneiro. (2021). A discriminação algorítmica e as novas perspectivas sobre o tratamento de dados pessoais sensíveis. In: REQUIÃO, Maurício (Org.). *Proteção de dados pessoais: novas perspectivas*. Salvador: Editora da Universidade Federal da Bahia.

DA EMPOLI, Giuliano. (2019). *Os engenheiros do caos*. Rio de Janeiro: Vestígio.

ELESBÃO, Ana Clara Santos; SANTOS, Jádia Larissa Timm dos; MEDINA, Roberta da Silva. (2020). Quando as máscaras (do reconhecimento facial) caírem, será um grande carnaval. In: SABARIEGO, Jesús; AMARAL, Augusto Jobim do; SALLES, Eduardo Baldissera Carvalho (Orgs.). *Algoritarismo*. 1. ed. São Paulo. Editora Tirant lo Blanch, p. 247-259.

FOUCAULT, Michel. (1999). *Em defesa da sociedade*. São Paulo: Martins Fontes.

FRANCISCO, Pedro Augusto P.; HUREL, Louise Marie; RIELLI, Mariana Marques. (2020, jun.). Regulação do reconhecimento facial no setor público: avaliação de experiências internacionais. Instituto Igarapé + Data Privacy Brasil Research, Disponível em: https://igarape.org.br/wp-content/uploads/2020/06/2020-06-09-Regula%C3%A7%C3%A3o-do-reconhecimento-facial-no-setor-p%C3%BAblico.pdf Acesso em: 13 out. 2023.

FRAZÃO, Ana. (2021, ago.). Discriminação algorítmica: por que os algoritmos preocupam quando acertam e quando erram? Jota, Disponível em: https://www.jota.info/opiniao-e-analise/colunas/constituicao-empresa-e-mercado/discriminacao-algoritmica-por-que-algoritmos-preocupam-quando-acertam-e-erram-04082021 Acesso em: 07 set. 2023.

GARRET, Filipi. ROBOS (BOTS). Disponível em: https://www.techtudo.com.br/noticias/2018/07/o-que-e-bot-conheca-os-robos-que-estao-dominando-a-internet.ghtml. Acesso em: 21 jan. 2024.

GERVASONI, Tássia A.; DIAS, Felipe V. (2023, set-dez). Violações de Direitos Humanos pelas Big Techs: contribuições do pensamento decolonial e de uma leitura criminológica do dano social. *Revista de Garantias e Direitos Fundamentais*, Vitória, *24*(03), p. 137-173.

GOMES, Sheley; MOURA, Iara. (2022). De Oakland ao Jacarezinho: os sistemas de reconhecimento facial precisam ser banidos. Revista Capital. Disponível em: https://www.cartacapital.com.br/blogs/de-oakland-ao-jacarezinho-os-sistemas-de-reconhecimento-facial-precisam-ser-banidos Acesso em: 21 jan. 2024.

LEMOS. André. (2021, maio-ago.). Dataficação da vida. Revista Civitas. Disponível em: http://dx.doi.org/10.15448/1984-7289.2021.2.39638 Acesso em: 09 jan. 2024.

MAGNO, Madja Elayne da Silva Penha; BEZERRA, Josenildo Soares. (2020, ago./dez.). Vigilância negra: o dispositivo de reconhecimento facial e disciplinaridade dos corpos. *Revista Novos Olhares*, *09*(2). Disponível em: https://www.revistas.usp.br/novosolhares/article/view/165698 . Acesso em: 30 jan. 2024.

MELO, Paulo Victor; SERRA, Paulo. (2022). Tecnologia de Reconhecimento Facial e Segurança Pública nas Capitais Brasileiras: Apontamentos e Problematizações. *Comunicação e Sociedade*, *42*, p. 205-220. Disponível em: https://journals.openedition.org/cs/8111 Acesso em: 23 jul. 2024.

NUNES, Fernanda Todesco *et al*. (2016). Um estudo sobre técnicas de biometria baseadas

em padrões faciais e sua utilização na Segurança Pública. *In*: SPANHOL, Fernando J.; LUNARDI, Giovani M.; SOUZA, Márcio Vieira de (org.) Tecnologias da Informação e Comunicação na Segurança Pública e Direitos Humanos. Coleção Mídia, Educação, Inovação e Conhecimento, *2*, Editora Edgard Blücher Ltda., p. 113-132.

NOBLE, Safiya. (2021). Algoritmos da opressão. Santo André: Rua do Sabão.

NUNES, Pablo. (2019). Novas ferramentas, velhas práticas: reconhecimento facial e policiamento no Brasil. *In*: REDE de Observatório de Segurança. *Retratos da violência: cinco meses de monitoramento, análise e descobertas*. Centro de Estudos em Segurança e Cidadania.

NUNES, Pablo. (2021, jan. 02). O algoritmo e racismo nosso de cada dia. *Revista Piauí*, Disponível em: https://piaui.folha.uol.com.br/o-algoritmo-e-racismo-nosso-de-cada-dia/ Acesso em: 23 jul. 2024.

OTEGUI, C. A. et al. (2006). *Proyecto Aguará – Reconocimiento de Caras*. Montevideo: Facultad de Ingeniería Universidad de la República.

OTEGUI, C. A. et al. (2021, jan. 02). O algoritmo e racismo nosso de cada dia. Folha de São Paulo [on line], São Paulo, Questões de vida digital. Disponível em: https://piaui.folha.uol.com.br/o-algoritmo-e-racismo-nosso-de-cada-dia Acesso em: 07 jan. 2024.

POELL, Thomas; NIEBORG, David; DIJCK, José van. Plataformização. (2020, jan. abr.). Plataformização. *Fronteiras - estudos midiáticos*, *22*(1). Disponível em: https://revistas.unisinos.br/index.php/fronteiras/article/view/fem.2020.221.01 Acesso em: 23 jul. 2024.

ROCHA, Jannotti da; PORTO, Lorena Vasconcelos; ABAURRE, Helena Emerick. (2020). Discriminação algorítmica no trabalho digital. *Revista de Direitos Humanos e Desenvolvimento Social*, *1*, e 205201. Disponível em: https://seer.sis.puc-campinas.edu.br/direitoshumanos/article/view/5201/3164. Acesso em: 07 dez. 2023.

ROLA, Eulálio do Carmo da Silva. (2022). Os principais contributos da inteligência artificial para o processamento de imagens digitais a utilizar na Segurança Pública. 2022, 146 f. Dissertação (Mestrado em Segurança e Justiça) – Universidade Lusíada, Lisboa.

ROSA, Pablo Ornelas; AMARAL, Augusto Jobim do; NEMER, David Baião. (2023). Datapolítica, governamentalidade algorítmica e a virada digital: uma genealogia da modulação comportamental através das plataformas digitais. *Revista Eletrônica do Curso de Direito da UFSM*, *18*(03). Disponível em: https://periodicos.ufsm.br/revistadireito/article/view/85510 Acesso em: 23 jul. 2024.

ROUVROY, Antoinette; BERNS, Thomas. (2015). Governamentalidade algorítmica e perspectivas de emancipação: o díspar como condição de individuação pela relação? *Revista Eco Pós. Rio de Janeiro*, *18*(2), p. 36-56. Disponível em: https://doi.org/10.29146/eco-pos.v18i2.2662 Acesso em: 18 nov. 2023.

SILVA, Tarcízio. (2020). Visão computacional e racismo algorítmico: branquitude e opacidade no aprendizado de máquina. *Revista ABPN*, *12*, p. 428-448. DOI: 10.31418/2177-2770.2020. Disponível em: https://abpnrevista.org.br/site/article/view/744/774 Acesso em: 07 set. 2023

SILVEIRA, Sérgio Amadeu. (2016). Governo dos algoritmos. *Revista de Políticas Públicas*, *02*(01), p. 267-281. Disponível em: https://edisciplinas.usp.br/pluginfile.php/4452794/mod_resource/content/1/S%C3%A9rgio%20Amadeu%20SILVEIRA%20%20Governo%20dos%20Algoritmos.pdf Acesso em: 26 jan. 2024.

SILVEIRA, Sérgio Amadeu. (2020). Sistemas algorítmicos, subordinação e colonialismo de de dados. In: SABARIEGO, Jesús; AMARAL, Augusto Jobim do; SALLES, Eduardo Baldissera Carvalho (organizadores). *Algoritarismo*. 1. ed. São Paulo: Editora Tirant lo Blanch, p. 158-1709.

SOARES, Marcelo Negri et al. (2022). Inteligência artificial e discriminação: um panorama sobre a antagonização entre exclusão e o Estado Democrático de Direito Brasileiro à luz dos direitos da personalidade. *Revista Direitos Sociais e Políticas Públicas (UNIFAFIBE)*, *10*(2), p.

567-597. Disponível em: https://portal.unifafibe.com.br/revista/index.php/direitos-sociais-politicas-pub/article/view/1311 Acesso em: 07 jan. 2024.

SRNICEK, Nick. (2018). *Capitalismo de plataforma*. Buenos Aires: Ed. Caja Negra.

SILVA, Tarcízio et al. (2020, jun. 5). Apis de visão computacional: Investigando mediações algorítmicas a partir de estudo de bancos de imagens. *Logos,* 1(27), pp. 25-54. Disponível em: https://www.e-publicacoes.uerj.br/logos/article/view/51523/33928 *Acesso em: 26 jul. 2024.*

TAUTE, Fabian. (2020 fev. 7). Reconhecimento Facial e suas controvérsias: O reconhecimento facial não só traz a possibilidade de instaurar uma vigilância em massa, mas também contém uma tendência preconceituosa contra certos grupos de nossas sociedades – com as mulheres negras sendo as mais afetadas. Heinrich Böll Stiftung, Rio de Janeiro. Disponível em: https://br.boell.org/pt-br/2020/02/05/reconhecimento-facial-e-suas-controversias Acesso em: 07 set. 2023.

TRASLAVIÑA, C. M. G. (2007). Introducción a la biometría. Disponível em: https://www.academia.edu/9374109/Introducci%C3%B3n_a_la_biometr%C3%ADa Acesso em: 17 nov. 2023.

VAIDHYANATHAN, Siva. (2011). A Googlelização de tudo. São Paulo: Cultrix.

ZUBOFF, Shoshana. (2020). *A era do capitalismo de vigilância: Luta por futuro humano na nova fronteira de poder.* Rio de Janeiro: Ed. Intrínseca.

PIONEER
ACADEMIC PUBLISHING LIMITED

Studies in Law and Justice
ISSN 2958-0382
www.pioneerpublisher.com/slj
Volume 4 Number 5 October 2025

The Dilemma and Solutions in Applying Criminal Law to Generative AI Fraud Crimes: The Case of Sora

Futian Zhang[1]

[1] Law School of Zhejiang University of Finance & Economics, Zhejiang, China
Correspondence: Futian Zhang, Law School of Zhejiang University of Finance & Economics, Zhejiang, China.

doi:10.56397/SLJ.2025.10.09

Abstract

The rapid advancement of artificial intelligence systems such as Sora has brought unprecedented convenience to humanity, yet simultaneously given rise to novel forms of fraud. The emergence of such AI systems does not alter the fundamental reality that machines lack autonomous consciousness; consequently, artificial intelligence cannot constitute the principal agent in fraud offences. Fraud crimes utilising generative AI like Sora lack the requisite intent for fraud, necessitating adherence to the "disposition consciousness requirement theory". Liability for fraud crimes should consider multiple parties—producers, users, intermediary institutions—based on their specific involvement. Addressing responsibility in AI-facilitated fraud crimes urgently requires establishing reasonable duties of care, refining relevant criminal charges and judicial interpretations for AI fraud, and constructing multi-tiered, multi-stakeholder regulatory mechanisms.

Keywords: criminal subject, criminal liability, fraud offences, disposition awareness, duty of care

1. Problem Statement: Development Principles and Challenges of Generative AI Models like Sora

On 15 February 2024, OpenAI unveiled the Sora artificial intelligence model to the public. Its name, derived from the Japanese word for "sky," signifies boundless creative potential. Developed from the text-to-image generation model DALL-E, Sora represents a significant breakthrough in video generation technology. It can produce videos lasting up to one minute based on instructions and scene prompts, recreating richly detailed real-world scenarios with heightened vividness and realism. The development of AI systems like Sora draws initial inspiration from the human brain. Computer scientists, guided by principles of neural architecture, continually refine deep learning algorithms, harness GPU computing power, and leverage vast training datasets to achieve AI capabilities in image recognition, speech recognition, and artistic creation. (Xianfeng Gu, 2016) Should fraudsters exploit AI systems like Sora to generate vast quantities of phishing emails, deceptive links, or "manipulative" videos based on commands, thereby inducing or reinforcing cognitive errors in victims, such programme-targeted or programme-enabled fraudulent activities may constitute fraud offences.

Artificial intelligence possesses deep learning capabilities and an ever-expanding scope for learning. As training periods lengthen and models continually refine, humanity may lose its intellectual edge, instead becoming deceived or manipulated by AI. (Xianquan Liu, 2023) Consequently, perpetrators need only utilise AI systems like Sora to gather others' videos or photographs for training and recognition, thereby enabling hyper-realistic "face-swapping fraud". Fraud syndicates, through organised planning, execute "highly efficient" and precise scams. (Qiangqiang Liu, 2024) Consequently, cyber fraud often spans multiple jurisdictions, presenting challenges in criminal law application, case determination, and evidence collection. Moreover, for fraud charges to hold, a causal link must exist between the perpetrator's acquisition of property and the victim's act of delivery or disposal. This demonstrates how AI intervention challenges existing legislation.

Sora can generate realistic and imaginative scenes from textual prompts, rendering it nearly indistinguishable from real-world footage to the human eye. Generative AI systems like Sora may undermine the voluntary nature of property disposal decisions and even deprive victims of the awareness that they have been deceived. Consequently, the author contends that further clarification is needed regarding the distinction between voluntary and involuntary property disposal in theft and fraud offences. Some scholars argue that in fraud cases, the voluntary nature of property disposal should be the sole consideration, regarding the specific motive behind property disposal is deemed irrelevant. (Gang Wang, 2014) Whether an offender exploiting an AI-induced error of perception or deliberately creating programme vulnerabilities can be classified as fraud remains contentious.

Existing "AI face-swapping technology" generates voice and video content for telecommunications fraud. In August 2024, overseas criminal syndicates employed deep integration of facial and video data. For instance, in the "AI Musk case," the esteemed entrepreneur promoted a fund investment opportunity promising rapid returns, with lip-syncing, voice, and body language highly matching the celebrity's. The advertisement even incorporated his signature South African accent, convincing the elderly victim. Driven by greed and curiosity, he contacted the foreign exchange company behind the ad, ultimately losing over $690,000 in his investment. Evidently, the emergence of generative AI systems like Sora will significantly impact the application of criminal law to fraud offences. (Shuiling Liu, Wenkai Dong & Wenkai Dong, 2024) While perpetrators may utilise videos generated by such AI to commit fraud, thereby constituting the offence of fraud, the legal characterisation of AI's role in providing and utilising criminal tools remains complex. (Bin Yuan, Liming Xue & Liming Xue, 2024) For instance, the act of disposing of property — an element not explicitly defined in fraud offences — is crucial in distinguishing fraud from theft.

In practical terms, the emergence of AI systems like Sora substantially elevates both the risk of ordinary individuals falling victim to fraud and the difficulty of detecting such scams. The lowered threshold for criminal activity, coupled with easier identity theft, means increasing numbers of ordinary people may become potential victims. Current criminal law proves inadequate for reasonably allocating liability and imposing criminal penalties on providers, producers, or users of generative AI services. This paper will analyse the potential fraud risks posed by generative AI systems like Sora and propose countermeasures to explore legal approaches for combating AI-facilitated fraud in China.

2. The Dilemma in Recognising Fraud Crimes Involving Generative AI Systems Like Sora and the Desired Position

2.1 Generative AI Systems Should Not Be Deemed Subjects of Fraud Offences

In this era of rapid AI advancement, criminal law must evolve with both foresight and reasonableness. Some scholars argue that generative AI systems like Sora possess the self-control and discernment required for fraud, thereby establishing AI as a subject of criminal liability. Humans are no longer the sole possessors of intelligence; the "intelligence" created by humanity now surpasses human capabilities. AI is categorised into strong AI, weak AI, and super AI based on whether intelligent robots can autonomously execute actions beyond human design and programming. (Xianquan Liu, 2019) Weak AI cannot constitute a criminal subject, whereas strong AI, possessing self-control and discernment, can independently make decisions and commit acts severely harmful to society,

thereby qualifying as a criminal subject. (Xianquan Liu, 2019) Superintelligence represents an existence surpassing the human brain in all aspects, transcending human cognitive limitations. Following a technological "singularity" breakthrough, it may develop free consciousness—a prospect warranting serious consideration (Liangfang Ye, 2019).

The author contends (Bencan Li, 2020) that law should neither seek to nor be suited for establishing artificial intelligence as a criminal subject. Firstly, the dividing line between affirmative and negative theories on criminal subjecthood lies in whether AI can generate independent consciousness. The latest generative AI systems like Sora have not demonstrated cases of self-awareness independent of programmatic control, and the "singularity theory" championed by affirmative scholars lacks scientific substantiation.

The emergence of human autonomous consciousness differs fundamentally from generative AI's comprehension of human language and deep learning principles. Current AI learning techniques remain confined within established frameworks, with progress limited to expanded data scales, enhanced precision, and improved integration. The prevailing approach involves simulating human neural network operation (Chong Wang & Puyu Dong, 2020), thus precluding the miraculous emergence of self-control or self-awareness. Human autonomous consciousness derives from a special genetic trait acquired through evolution, endowing humans with the capacity to filter endogenous information.

From the perspective of legal subjectivity and self-interest, artificial intelligence cannot be treated as a legal person in the same manner as a corporate entity. The original intent behind humanity's creation of AI was to study an "object" more beneficial to itself, rather than to create a "new life form". (Ji Yang, 2019) As a tool, AI lacks volition and free will in its actions; consequently, harm caused by it cannot be attributed to it. (Weipan Si, 2020) From a legal-philosophical perspective, as a legal concept, the object of punishment "is not a purely objective entity, but a socially constructed phenomenon—a value fact." (Mu Wang, 2018) Philosophically, a subject of criminal law must possess the essential characteristics of practicality and sociality. Practicality requires the subject to engage in conscious activity, while sociality demands the subject's capacity to interact within specific social relationships (Libin Wang, 2019).

Some scholars contend that in the future era of super-intelligent artificial intelligence, AI could become subjects within legal relationships, subject to legal regulation and constraints. Similar to the concept of a legal person, the legal consequences of its actions would be borne by the AI itself, possessing independent capacity for action, rights, and liability. (Yunfeng Wu, 2018) However, the author believes that the developmental trajectory of artificial intelligence is to serve and benefit humanity. A legal person can be regarded as a collective of multiple individuals, and determining the liability of a legal person essentially involves the rights and obligations between natural persons. The affirmative view evidently overlooks the fact that legal persons themselves possess no free will, whereas the governing bodies of legal persons comprise real individuals who naturally possess free will. In crimes committed by intelligent robots, no natural persons are present—how then can human free will be invoked (Hongbing Chen, 2021)?

If criminal law seeks to penalise artificial intelligence infringing upon others' rights by imposing human-centric punishments—such as fines or restrictions on liberty—these would fail to instil fear in cold machinery. Taking Professor Liu Xianquan's proposal for three special penal measures against intelligent robot crimes as an example: "deleting data," "modifying programmes," and "permanent destruction" could be applied to criminal artificial intelligence. From the perspective of penal objectives, such measures must functionally possess punitive attributes. From a utilitarian standpoint, they must achieve both general and specific deterrence. Intelligent robots, devoid of pain or moral sensibility, cannot possibly experience the so-called "punishment" as suffering.

2.2 Fraudulent Crimes Utilising Generative AI such as Sora Lack Fraudulent Intent

The fraudulent intent in traditional fraud offences refers to the perpetrator's purpose of unlawful appropriation, where they knowingly fabricate facts or conceal truths, thereby inducing others into erroneous beliefs and improper disposal of property, anticipating and desiring the resulting financial loss. Take

traditional telecommunications fraud as an example: fraud syndicates deceive victims through manipulative language or actions, with the fraudulent intent residing in the victim's potential reception of such manipulative information.

The challenge in establishing fraud when using generative AI like Sora lies in scenarios where others cannot receive the fraudulent information. The perpetrator possesses the intent to secretly steal another's property but lacks the traditional criminal intent. Take Alipay as an example: its user agreement stipulates that identity credentials are crucial for user verification and must be safeguarded. Any transaction using these credentials is deemed authorised by the account holder. However, most users merely register to complete the process without scrutinising every clause. Should fraudsters successfully obtain users' biometric information via generative AI systems like Sora, or collect biometric data from other platforms to synthesise profiles for Alipay breaches, such deception constitutes covert theft.

AI deepfake technology heightens fraud risks while complicating the establishment of fraudulent intent. Victims may lack awareness of receiving deceptive communications, thereby lacking the requisite intent for fraud. In February 2024, the Hong Kong branch of a British multinational corporation suffered HK$200 million losses after being deceived by AI-generated images and audio of its CEO, created by fraudsters using deepfake technology. The fraudsters employed this technique to fabricate executive team members, thereby gaining employees' trust. The core principle of such deepfake videos lies in algorithms like generative adversarial networks or convolutional neural networks, which transplant facial features from one subject onto another. Videos generated by AI systems such as and Sora are composed of sequential images forming dynamic footage. Thus, by altering facial features in each frame, multiple images can be synthesised into a fake video, enabling automated deep learning fraud. Guided by the lifelike imagery of the "executive team," employees routinely executed transfer tasks without recognising the fraudulent prompts. This demonstrates that victims may not consciously receive manipulative messages within specific scenarios, yet their personal information and assets are already compromised through AI's "stealthy extraction."

2.3 Generative AI Fraud Crimes like Sora Uphold the "Necessity of Disposition Awareness" Doctrine

To explore the concept of punishment awareness in fraud offences, one must first clarify the relationship between theft and fraud. There exists no relationship of imaginary concurrence between theft and fraud; acts not constituting theft cannot be classified as such merely because theft carries heavier penalties. Fundamentally, theft and fraud represent two distinct criminal offences (Wenhan Liu & Shixin Liu, 2023).

Both theft and fraud fall under the category of crimes involving the transfer of possession. Traditionally, possession is understood as the state of factual control over a specific object exercised by an individual based on subjective intent. In theft, the perpetrator surreptitiously transfers property held by the victim to their own possession against the victim's will. Fraud, however, is predicated on deceptive communication between the perpetrator and the victim. When customers scan codes to make payments at merchants, criminal syndicates exploit AI-enabled fraud to illegally obtain account details from payment platforms like Alipay and WeChat. Using AI, they generate numerous QR codes. Victims are tricked into clicking account verification links (deceptive links) to activate their accounts. The moment the victim's payment code appears, it is replaced with a fraudulent code, enabling account intrusion and unauthorised transactions. Some scholars contend this should be classified as theft, arguing that fraud requires the victim's dispositive act to unlawfully acquire property, whereas theft does not necessitate such consent. They contend that swapping QR codes involves no communication between customer and merchant; the customer objectively engages in a normal transaction without deception, thus lacking the dispositive act of transferring funds to the perpetrator. Consequently, it should be classified as theft. The author contends that such scenarios constitute fraud offences. The perpetrator's substitution deceives both the merchant and the customer. The customer disposes of property based on the merchant's instruction to "scan and pay". From a triangular fraud perspective, the deceived party is the customer, the victim is the merchant, and the merchant suffers property loss due to the customer's mistaken belief. This establishes a triangular fraud relationship.

Whether perpetrators employing generative AI like Sora possess the requisite dispositive intent for fraud remains subject to further debate. Fraud requires not only deceptive methods but also the victim's property transfer based on mistaken belief. When telecommunications fraud first emerged, judicial practice already recognised "unwitting delivery" as a relevant concept. Theories on dispositive intent in fraud offences can be categorised into the "necessity theory" and the "non-necessity theory". The necessity theory holds that the victim must subjectively recognise they are transferring property; without dispositive intent, fraud cannot be established. The non-necessity theory contends that dispositive intent is not essential to fraud; the offence is established by the objective transfer of property alone. Ryūichi Hirano advocates the non-necessity theory, asserting that when the object of fraud is property, "it suffices that there be a factual act of transferring possession; recognition of this is unnecessary, and unconscious delivery (disposition) is also permissible." Norihiko Nishida similarly contends: Fraud is established where possession of property or pecuniary advantage demonstrably transfers to the other party based on the defrauded person's intent. Excluding the most typical scenario—concealing the transferred object (the criminal target)—from fraud charges is inappropriate. Thus, unconscious dispositive acts should suffice to constitute the dispositive conduct required for this offence. "In fraud offences, the victim disposes of property while in a state of free will. Their consent to such disposal stems from the mistaken belief that the act will yield reasonable consideration, thereby prompting their decision to relinquish possession. Consequently, property loss in fraud offences arises only through the victim's 'voluntary cooperation'" (Lizhi Wang, 2015).

Conversely, if dispositive intent were disregarded, the scope of dispositive acts would be indefinitely expanded. For instance, in the Hong Kong branch case, had the fraudsters impersonated senior executives and routinely assigned tasks to employees for collecting funds—without using video calls—staff might tacitly accept this as normal work arrangements, thereby constituting an "omission" offence. Such tacit acquiescence lacks objective dispositive action. Thus, while the absence of objective dispositive conduct may not preclude conviction based on dispositive intent, cases lacking subjective dispositive intent fail to distinguish theft from fraud. This renders the argument formally illogical (Langtao Bai, 2017).

The author contends that in AI-facilitated fraud, victims may lack both the intent to dispose of assets and engage in self-defeating cooperation. Pre-Sora AI scams already mimicked human behaviour and linguistic patterns—such as forging users' writing styles to send fraudulent emails prompting clicks on malicious links for sensitive data disclosure; AI bots automatically interact to coax victims into revealing personal details or transferring funds; AI systems analyse victims' software usage habits to launch phishing or malware attacks at optimal moments. In these scenarios, deception stems from victims' unfamiliarity with software operation techniques and fraud tactics. Technological ignorance leads them to mistakenly believe clicking a link or granting software permissions poses no security risk, thus lacking the requisite intent to dispose of property.

The author contends that victims in AI-enabled fraud may possess a sense of financial agency, albeit one compromised by irrationality and complacency towards fraudulent techniques, leading to self-inflicted cooperation. With the advent of Sora, deepfakes simulate the real world by generating complex scenarios featuring multiple characters and scripted actions. As illustrated by the 2024 Hong Kong subsidiary fraud case, employees accepted video call invitations from fraudsters where the depicted colleague was indeed a company partner; In reality, the fraudsters employed deepfake technology to forge the executive team and gain the employees' trust. When victims are deceived by AI fraudsters like Sora and misled by deepfake technology, they develop the intent to dispose of their assets. Subjectively, they recognise they are transferring property, and objectively, they carry out the transfer. According to the "necessary intent theory of disposing of property," this constitutes the crime of fraud.

2.4 Liability for Generative AI Fraud Crimes Involving Sora and Similar Systems

Human criminal liability is enforced through deprivation of liberty, property, or even life. Applying such penalties to AI systems like Sora—depriving them of freedom or confining

them—would not induce remorse or repentance, especially given its lack of consciousness. The deterrent function of criminal law would be entirely negated. Ordering AI to pay compensation or terminating its existence is preposterous. Even interpreting termination as decommissioning would entail immense waste of human and material resources. The author contends that liability should instead be borne by the producers, users, or third-party institutions of AI systems like Sora, according to varying degrees of duty of care. By configuring obligations and criminal responsibility across different entities, we can achieve the dual objectives of preventing AI-related crimes and upholding ethical values in technology (Di Sun, 2022).

Firstly, when multiple parties—producers, users, and intermediary institutions—participate in fraud crimes, determining criminal liability requires case-by-case analysis. When producers deliberately design AI systems like Sora for fraudulent purposes, and users acquire such systems specifically for criminal exploitation, questions arise regarding the allocation of criminal liability between producers and users. Some scholars contend that AI producers are no different in essence from manufacturers of conventional goods, and thus producers should be held criminally liable under Articles 140 to 150 of the current Criminal Code. However, the author contends that such a blanket classification is overly simplistic and fails to account for the specific circumstances of each case.

Secondly, while producers develop AI like Sora to serve humanity, users deploy it for fraudulent crimes. Users should bear responsibility. If producers develop AI programmes containing vulnerabilities of which they remain unaware, and the user exploits this vulnerability to commit fraud, criminal liability should be borne separately by both the producer and the user. The producer should be held liable based on the extent of losses caused by the programme vulnerability. However, in practice, it is difficult to assess the losses attributable to programme vulnerabilities. Some scholars argue that as professional technicians, developers possess both theoretical analytical and practical operational capabilities during the AI development process, enabling them to make scientific judgements regarding the security risks of AI products. Therefore, they should bear a proportionately greater duty of care. This viewpoint also holds merit.

3. Pathways for Refining Criminal Law Regulation of Generative AI Fraud

3.1 Establishing Criminal Duty of Care

Establishing criminal-law duties of care for designers, manufacturers, and users of AI products is pivotal to addressing crimes where AI serves as the "substance" (Liangfang Ye, 2019). This approach also better resolves liability issues in AI fraud crimes like those involving Sora. However, academic debate persists regarding the source of such duties: some scholars advocate a dual "overall + individual" approach. This would involve establishing systemic boundaries for heightened duties on AI algorithm service developers at the collective level, while applying a "reasonable person" standard at the individual level to refine the assessment of heightened duties in specific cases, thereby balancing all stakeholders' interests. (Yingying Yang, 2024) Nevertheless, the author contends that imposing a heightened duty of care risks indefinitely expanding liability boundaries, potentially stifling the momentum of AI industry development. It also risks subjecting users' legitimate activities to unreasonable interference, thereby undermining the long-term sustainable growth of the digital economy.

Some scholars propose two potential pathways depending on the domain: one involves expanding the scope of criminal negligence by broadening the concept of negligence to hold human designers and users accountable, the other being a complete prohibition on AI usage in domains closely tied to significant personal and societal interests, thereby reducing the burden of duty of care on natural persons. However, the author contends that the second path lacks practicality. One cannot directly ban possibilities merely because "risks" exist, especially when employing criminal law to forcibly dissolve cutting-edge technologies, which contradicts the principle of restraint in criminal law and hinders technological advancement. The feasibility and rationality of the first approach also remain questionable, though its progressive merit lies in establishing a duty of care for natural persons, which at the very least serves as a cautionary measure (Chenceng Chu, 2018).

Academic discourse remains divided on

categorising duty of care obligations across different entities and scenarios. Some scholars propose tailoring obligations to the identity of the responsible party. Where developers or designers breach such duties, they should bear corresponding legal liability. Criminal offences may be prosecuted under Article 146 of the Criminal Law concerning the production or sale of products failing to meet safety standards. Producers and sellers retain a degree of control over AI systems. Beyond conducting preliminary criminal risk assessments, they must also inform users of potential hazards, such as potential programme vulnerabilities during Sora's operation or guidelines for appropriate video usage. Users' duty of care should further encompass regular maintenance checks to ensure the AI functions correctly. (Renqiang Sun & Daoyuan Wang, 2023) Some scholars propose further categorising the duty of care into intentional, negligent, and accidental scenarios. Intentional cases involve a clear failure to exercise due diligence, potentially including deliberate disruption of legal order. Negligent scenarios may arise when users, unfamiliar with Sora's command operations during initial use, commit errors due to failure to meet the standard of care expected of a reasonable person. Such negligence may be deemed a breach of the duty of care, leading to criminal liability under the principle of legality, and may lead to operational errors during initial use due to unfamiliarity with command interfaces. Failure to meet the standard of care expected of a reasonable person would constitute a breach of duty, potentially warranting criminal liability under the principle of legality. In unforeseeable circumstances, where the involved party has demonstrated reasonable diligence and provided evidence of due diligence, higher standards of care cannot be imposed.

Regarding the establishment of criminal due diligence obligations, the author summarises as follows: Firstly, the scope of due diligence obligations must not exceed human control or cognitive capacity; that is, obligations should not be excessively stringent, lest they unduly burden all parties. Secondly, from a practical perspective, establishing due diligence obligations for AI developers necessitates concurrently establishing technical standards for artificial intelligence, constraining designers' methodologies to thereby limit functional choices. Third, the duty of care should be confirmed according to different subjects and circumstances (intent, negligence, and unforeseeable events). Where the involved subject has demonstrated reasonable diligence and provided evidence thereof, a higher standard of care cannot be demanded. Fourth, the restraint and forward-looking nature of criminal law must be consistently upheld. The existence of "risk" should not directly preclude the possibility of AI advancement in fields closely tied to significant individual and societal interests.

3.2 Establish New Offences for AI-Related Fraud and Refine Relevant Judicial Interpretations

With the advent of artificial intelligence systems such as Sora, both the exploitation of AI for fraudulent criminal activities and the manipulation of AI systems to deceive others for personal gain represent formal innovations. If analysed solely within the framework of traditional fraud offences, existing criminal law provisions prove inadequate. These shortcomings may be addressed through legislative refinement or by extending the interpretation of current criminal statutes.

Some scholars propose refining the current Criminal Code, such as amending "fraud" to "fraud or artificial intelligence fraud". The author contends such modifications would be overly cumbersome. Moreover, the continuous advancement of artificial intelligence will inevitably give rise to numerous new issues. It would be preferable to follow the model of the existing "illegal use of information networks" offence and create a dedicated criminal charge specifically for artificial intelligence crimes. Some scholars propose introducing a "crime of unlawful utilisation of artificial intelligence." Such an offence could encompass multiple scenarios involving traditional crimes, enumerating AI-related criminal acts beyond mere fraud. Under this single offence, specific duties of care would be stipulated for multiple parties, including AI producers, users, and intermediaries. However, the author contends that this approach fails to cover cases of negligence or unforeseen incidents where perpetrators cause major safety incidents due to inadequate AI management obligations, necessitating further refinement. Some scholars have addressed this deficiency by proposing the establishment of a "crime of major artificial intelligence safety incidents." This offence

would be predicated on the perpetrator's violation of AI management obligations. Given the widespread deployment of AI in public settings, designers, producers, and users who fail to comply with stringent regulations or neglect relevant duties of care could be penalised under this provision. Some scholars contend that artificial intelligence possesses autonomous learning capabilities and the potential to operate beyond human control. Consequently, determining the negligence of relevant human entities requires case-by-case analysis. They propose introducing a new offence of "negligent harm caused by the research, development, production, or sale of artificial intelligence", imposing prior duties of care on producers and sellers, with liability for negligence arising from failure to fulfil these duties resulting in actual harm. The user's negligence should be determined comprehensively based on their capacity to foresee consequences and avoid them. (Luyao Ma, 2023) The author considers such a framework reasonable, establishing the duty of care for all parties as the statutory basis for pursuing liability for actual harm caused by artificial intelligence infringing upon legal interests.

Considering two scenarios—malicious exploitation of AI for criminal purposes and AI operating beyond human control causing negligence—the author proposes introducing two additional offences. Firstly, establish the "Offence of Illegal Utilisation of Artificial Intelligence", with sub-offences including illegal AI-assisted fraud and illegal AI-assisted intellectual property infringement. For instance, using AI like Sora to generate videos depicting a child's abduction, simulating voices and scenarios to defraud parents of their assets, would constitute "AI-assisted fraud." The intent and dispositive awareness required for fraud under this offence should be defined, appropriately expanded to account for the unique role of AI involvement, with liability allocation and supervisory responsibilities clarified in judicial interpretations. Second, introduce the offence of "AI liability accidents", distinct from the existing "major liability accident offence" in judicial practice. The latter requires violations occurring "during production or operations" and "breaches of relevant safety management regulations". This new offence would apply when reasonable duty of care is breached during production, use, or distribution, resulting in AI operating beyond human control to commit crimes causing tangible harm.

3.3 Establishing a Tiered and Categorised Regulatory Framework

In July 2024, the Decision of the Central Committee of the Communist Party of China on Further Comprehensively Deepening Reforms and Advancing Chinese-Style Modernisation proposed establishing an "artificial intelligence safety supervision system" to refine the governance framework for AI. Current oversight of AI developers remains unstandardised, undermining public interests and hindering technological advancement. Legal provisions must therefore define the due diligence and management obligations of various stakeholders to accelerate the regulation of generative AI.

Considering the forward-looking nature of criminal law, none of China's current criminal penalties are applicable to strong artificial intelligence robots. The nation's penal system urgently requires refinement. Furthermore, given that artificial intelligence cannot be punished, a tiered and categorised regulatory system must be established. Some scholars advocate that unified AI legislation should adhere to an inclusive and prudent regulatory philosophy, thereby establishing a "transition period" for the introduction of AI-related laws, policies, and standards. This could be achieved by introducing a regulatory sandbox system, providing a controlled environment for pilot testing. The author concurs with the concept of risk-based classification and differentiated regulation, advocating for case-by-case analysis to implement tiered, categorised, and differentiated oversight of AI applications (Hualin Song, 2024).

The European Union has already legislated for artificial intelligence, mandating the establishment of an AI regulatory sandbox system. China could draw upon this concept by testing artificial intelligence systems to predict the risk of fraud crimes potentially committed by AI systems such as Sora. Should an AI system demonstrate susceptibility to being "deceived" or "exploited" during testing, measures should be taken to mitigate security risks and enhance the accountability of development teams.

Generative AI systems like Sora are built upon typical black-box algorithmic models,

necessitating the development of optimised oversight frameworks such as anti-deepfake models. Anti-deepfake algorithms function as supervisors, identifying the fraudulent technique of "deepfakes" through keyword filtering, visual analysis, and prompt selection. Such systems can detect AI-generated face swaps by analysing varying degrees of facial distortion, or abstract facial expression movements into fundamental deformation units by observing and calculating micro-expressions and facial state analysis. This enables the identification of malicious AI-driven fraud while simultaneously detecting keywords and sensitive terms within AI conversations. By detecting criminal activity and linking directly to Chinese public security systems, timely alerts and reporting are facilitated, significantly reducing the incidence of AI-enabled fraud.

4. Conclusion

In October 2023, President Xi Jinping proposed the Global Initiative on AI Governance, advocating a shared consensus centred on humanity and the benevolent application of intelligence. This initiative promotes values of equality, mutual benefit, and respect for human rights, offering constructive solutions to widely debated issues concerning AI development and governance.

German scholar Ulrich Beck introduced the concept of the "risk society" in his work The Risk Society, Legislation invariably lags behind societal development. AI-enabled fraud crimes within this risk society not only challenge traditional criminal law theories but also impose entirely new demands upon the legal system. "Do not forget why you started out because you have gone too far." The emergence of AI fraud crimes prompts fresh reflection on the limits of criminal law intervention. The specific application of criminal law to regulating generative AI must strike a balance: neither too broad, thereby condoning disorderly development, nor too stringent, thereby stifling innovation in the generative AI market.

On one hand, criminal law must effectively regulate criminal conduct to uphold social order and safeguard citizens' rights; on the other, excessive legal intervention risks stifling technological innovation and advancement. Regarding liability attribution in AI-facilitated fraud, establishing reasonable tiered duties of care and constructing multi-level, multi-stakeholder regulatory mechanisms are essential. These measures not only concern the legal status of AI but also test the legal system's adaptability to emerging technologies.

In summary, the application of criminal law to AI-facilitated fraud crimes involves multiple dimensions, necessitating deep integration of legal, technological, and ethical considerations to establish a more comprehensive and effective legal framework that safeguards societal stability and citizens' rights. In judicial practice, multifaceted considerations must be integrated—including the application of fraud offences, personal information protection, technical forensics, complicity theories, balancing criminal intervention with technological innovation, the adaptability of criminal law, attribution of criminal liability, and ethical and legal challenges—to achieve precise targeting and effective regulation of AI-enabled fraud crimes.

References

Bencan Li. (2020). Criminal Liability of Natural Persons, Corporations, and Robots. *Contemporary Law, 34*(03), 99-109.

Bin Yuan, Liming Xue & Liming Xue. (2024). Research on Criminal Regulation of Generative Artificial Intelligence. *Hebei Law Review*, *42*(02), 140-159.

Chenceng Chu. (2018). The Direction of Criminal Liability Attribution in the AI Era: A Discussion Centered on the Attribution Gap of Negligence. *East China Journal of Law*, (03), 27-37.

Chong Wang & Puyu Dong. (2020). Re-examining Criminal Liability Subjects in the Era of Artificial Intelligence. *Guangxi Social Sciences*, (12), 118-125.

Di Sun. (2022). The Case Against Criminal Subject Status for Artificial Intelligence Entities. *Political and Legal Forum*, (03), 40-50.

Gang Wang. (2014). Objective Elements of Fraud in German Criminal Law: Focusing on German Judicial Precedents. *Politics and Law*, (10), 33-54.

Hongbing Chen. (2021). The Denial of Criminal Subject Status for Artificial Intelligence and Its Practical Implications: A Commentary on the Debate Between "Anti-Intellectual Criticism" and "Pseudo-Criticism". *Social Science Quarterly*, (06), 92-98.

Hualin Song. (2024). Regulatory Structure Design in Artificial Intelligence Legislation. *Journal of East China University of Political Science and Law*, *27*(05), 6-20.

Ji Yang. (2019). The Criminal Liability System in the Era of Artificial Intelligence Does Not Require Reconstruction. *Comparative Law Research*, (04), 123-137.

Langtao Bai. (2017). On the "Intent to Dispose" in Fraud Crimes. *East China Journal of Law*, (02), 97-106.

Liangfang Ye. (2019). Is Artificial Intelligence a Qualified Subject of Criminal Liability? *Global Legal Review*, *41*(04), 67-82.

Liangfang Ye. (2019). Is Artificial Intelligence a Qualified Subject of Criminal Responsibility? *Global Legal Review*, *41*(04), 67-82.

Libin Wang. (2019). Research on Criminal Liability Issues of Weak Artificial Intelligence. *Hunan Social Sciences*, (04), 57-63.

Lizhi Wang. (2015). The Necessity of "Disposition Awareness" in Establishing Fraud Charges: A Case Study of Fraudulent Acquisition Involving "Unwitting Delivery". *Forum on Politics and Law*, *33*(01), 119-131.

Luyao Ma. (2023). Research on AI Criminal Legislation Guided by Systems Thinking. *Juvenile Crime Issues*, (02), 20-32.

Mu Wang. (2018). On the Concept of Punishment: From "Essence" to "Significance". *Contemporary Law, 32*(02), 55-62.

Qiangqiang Liu. (2024). Governance Dilemmas and Countermeasures for AI Face-Swapping Fraud. *Cybersecurity Technology and Application*, (02), 160-162.

Renqiang Sun & Daoyuan Wang. (2023). Exploring Regulatory Pathways for Criminal Law in the AI Domain [C]// Beijing Criminological Research Association. Criminological Studies (Second Series). Criminal Court of Nanjing Jiangbei New Area People's Court; Nanjing Municipal Public Security Bureau, 139-147.

Shuiling Liu, Wenkai Dong & Wenkai Dong. (2024). Impact and Countermeasures of Generative AI like Sora on Judicial Determination of Fraud Crimes. *Juvenile Delinquency*, (03), 114-124.

Weipan Si. (2020). The Status and Allocation of Criminal Liability Subjects in Artificial Intelligence. *Chinese Journal of Applied Law*, (06), 172-186.

Wenhan Liu & Shixin Liu. (2023). Distinguishing Theft from Fraud in New Payment Methods. *Tianjin Law Review*, *39*(04), 41-51.

Xianfeng Gu. (2016). Historical Review and Current Development of Artificial Intelligence. *Nature China*, *38*(03), 157-166.

Xianquan Liu. (2019). A New Interpretation of the Concept of Conduct in Criminal Law in the Era of Artificial Intelligence. *Chinese Journal of Criminal Law*, (04), 60-72.

Xianquan Liu. (2019). A Response to the Denial of Criminal Liability Subject Status for Strongly Intelligent Robots. *Legal Review*, *37*(05), 113-121.

Xianquan Liu. (2023). Research on Criminal Liability Issues of Generative AI Systems like ChatGPT. *Modern Law Review*, *45*(04), 110-125.

Yingying Yang. (2024). The Reasonable Boundaries for Expanding the Duty of Care for Algorithm Recommendation Service Providers. *Lanzhou Journal*, (08), 88-103.

Yunfeng Wu. (2018). Dilemmas and Solutions in Applying Criminal Law to Property Crimes in the Era of Artificial Intelligence. *Law Science*, (05), 165-179.