Constitutional Safeguards Against Gender Bias in Artificial Intelligence: A Framework for Gender-Responsive AI Governance

Yayuk Whindari[1], Sheila Kusuma Wardhani Amnesti[2], Nurul Istiqomah[3], Siti Zulaichah[4]
[1,2,3,4] UIN Maulana Malik Ibrahim Malang, Indonesia

Abstract:

Artificial intelligence (AI) has become an integral part of decision-making in both public and private sectors, offering significant benefits while simultaneously raising concerns over algorithmic discrimination. Gender bias embedded in AI systems may produce unequal outcomes in employment, healthcare, finance, law enforcement, and public services, thereby threatening constitutional principles of equality, non-discrimination, human dignity, and due process. Although existing studies have extensively discussed AI ethics and regulatory governance, limited attention has been given to the constitutional dimensions of gender-biased AI. This article addresses this gap by examining how constitutional safeguards should be reconstructed within the framework of digital constitutionalism to respond to gender bias in AI.

This study employs normative legal research using statutory, conceptual, and comparative approaches by analyzing constitutional provisions, international human rights instruments, European AI regulatory frameworks, and contemporary scholarship on digital constitutionalism. The study finds that gender bias in AI constitutes a constitutional challenge because algorithmic systems increasingly exercise public-like power without adequate constitutional oversight. It argues that constitutional protection should incorporate the core principles of trustworthy AI, including human oversight, transparency, accountability, fairness and non-discrimination, privacy and data governance, technical robustness, and effective remedies, thereby strengthening digital constitutionalism as a framework for protecting fundamental rights in the age of artificial intelligence.

Keywords: Artificial Intelegence; Constitutional Rights; Digital Constitutionalism; Gender Bias

Background

Artificial intelligence (AI) has evolved from a computational technology into a critical infrastructure for decision-making across contemporary societies. Advances in machine learning, big data analytics, and computational capacity have enabled AI systems to perform functions that were previously dependent on human judgment, including classifying individuals, predicting behaviour, allocating resources, and supporting decisions in public administration, healthcare, education, employment, finance, and law enforcement. Consequently, AI no longer serves merely as a technological tool but increasingly operates as an institutional mechanism through

which opportunities, services, and public benefits are distributed.[1] This transformation has fundamentally altered the architecture of governance, shifting many decision-making processes from identifiable human actors toward algorithmic systems capable of operating at unprecedented speed, scale, and complexity.[2]

The growing reliance on AI has also reconfigured the exercise of power. Traditionally, decisions affecting legal rights and social interests were made by public officials or private actors whose authority was accompanied by legal responsibility and institutional accountability. Today, however, algorithmic systems increasingly participate in decisions that shape access to employment, healthcare, financial services, education, and government assistance.[3] This phenomenon has been described as algorithmic governance, in which computational systems perform functions traditionally exercised through human institutions. Although algorithms do not formally possess governmental authority, their practical influence over the distribution of rights, opportunities, and public services is often comparable to administrative decision-making. At the same time, algorithmic decisions are frequently characterised by opacity and limited explainability, making it difficult for affected individuals to understand, challenge, or seek remedies against automated outcomes.[4] As a result, the rapid expansion of AI raises concerns that extend beyond technological innovation to broader questions of transparency, accountability, democratic legitimacy, and the protection of fundamental rights.

Despite the widespread perception that AI produces objective and neutral decisions, growing evidence demonstrates that algorithmic systems frequently reproduce existing social inequalities.[5] Rather than creating discrimination independently, AI learns from historical datasets that often contain embedded patterns of bias, thereby generating unequal outcomes through seemingly objective computational processes.[6] This phenomenon, commonly referred to as *algorithmic bias*, illustrates that statistical optimisation does not necessarily produce substantive fairness. Bias may emerge from unrepresentative datasets, discriminatory historical records, model design, or deployment contexts, allowing structural inequalities to be reproduced at a much larger scale through automated decision-making. Consequently, algorithmic bias has become one of the most significant legal and governance challenges accompanying the widespread adoption of artificial intelligence.

---

[1] Laurence Diver, "Karen Yeung and Martin Lodge (Eds) Algorithmic Regulation Reviewed by Laurence Diver," *Prometheus* 37, no. 4 (2021), https://doi.org/10.13169/prometheus.37.4.0387.
[2] Pasquale, Frank. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, MA and London, England: Harvard University Press, 2015. https://doi.org/10.4159/harvard.9780674736061
[3] Karen Yeung, "Algorithmic Regulation: A Critical Interrogation," *Regulation & Governance* 12, no. 4 (2018): 505–23, https://doi.org/10.1111/rego.12158.
[4] Jenna Burrell, "How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms," *Big Data & Society* 3, no. 1 (2016): 2053951715622512, https://doi.org/10.1177/2053951715622512.
[5] Aylin Caliskan et al., "Semantics Derived Automatically from Language Corpora Contain Human-like Biases," *Science* 356, no. 6334 (2017): 183–86, https://doi.org/10.1126/science.aal4230.
[6] Andrew D. Selbst et al., "Fairness and Abstraction in Sociotechnical Systems," *Proceedings of the Conference on Fairness, Accountability, and Transparency* (New York, NY, USA), FAT* '19, January 29, 2019, 59–68, https://doi.org/10.1145/3287560.3287598.

Among the various forms of algorithmic discrimination, gender bias has received increasing international attention because of its persistent and cross-sectoral impact. Empirical studies have shown that AI systems may disadvantage women in employment recruitment, facial recognition, healthcare, financial services, and digital platforms. Amazon's experimental recruitment algorithm, for instance, systematically downgraded applications associated with women after learning from historically male-dominated employment data, while facial recognition technologies have repeatedly demonstrated lower accuracy in identifying women, particularly women with darker skin tones.[7] Similar disparities have also been identified in healthcare algorithms that underestimate women's medical risks due to datasets developed predominantly from male-centred clinical information.[8] These examples demonstrate that gender bias is not an isolated technical malfunction but a structural consequence of algorithmic systems trained on historically unequal social realities.

The implications of gender-biased AI extend far beyond unequal treatment in individual cases. As algorithmic systems increasingly determine access to employment, healthcare, education, financial resources, and public services, discriminatory outcomes directly affect the enjoyment of constitutionally protected rights. Unlike conventional discriminatory decisions made by identifiable human actors, algorithmic decisions are often opaque, difficult to explain, and challenging to contest through existing legal mechanisms. Consequently, gender bias threatens not only the principles of equality and non-discrimination but also broader constitutional guarantees, including human dignity, due process, procedural fairness, and access to effective remedies.[9] These developments indicate that gender bias in AI should no longer be viewed solely as a technical or ethical concern. Rather, it represents a constitutional challenge arising from the growing exercise of algorithmic power over fundamental rights.

Nevertheless, existing scholarship continues to examine gender bias primarily through the perspectives of computer science, AI ethics, and regulatory governance, with proposed solutions largely focusing on improving datasets, reducing statistical bias, and strengthening algorithmic fairness.[10] While these approaches remain important, they do not fully address how constitutional law should respond to algorithmic systems that increasingly exercise decision-making functions affecting fundamental rights. This gap suggests the need to move beyond technical governance towards a constitutional framework capable of ensuring that the exercise of algorithmic power remains consistent with the principles of equality, accountability, transparency, and the protection of human rights. It is within this context that digital constitutionalism provides an important normative foundation for reconstructing constitutional safeguards against gender bias in the age of artificial intelligence.

---

[7] Manish Raghavan, *The Societal Impacts of Algorithmic Decision-Making*, 1st ed. (ACM, 2023), https://doi.org/10.1145/3603195.
[8] Ziad Obermeyer et al., "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations," *Science* 366, no. 6464 (2019): 447–53, https://doi.org/10.1126/science.aax2342.
[9] Giovanni De Gregorio and Roxana Radu, "Digital Constitutionalism in the New Era of Internet Governance," *International Journal of Law and Information Technology* 30, no. 1 (2022): 68–87, https://doi.org/10.1093/ijlit/eaac004.
[10] Shira Mitchell et al., "Algorithmic Fairness: Choices, Assumptions, and Definitions," *Annual Review of Statistics and Its Application* 8, no. 1 (2021): 141–63, https://doi.org/10.1146/annurev-statistics-042720-125902.

Following the growing recognition that algorithmic discrimination threatens fundamental rights, governments and international organisations have introduced various regulatory initiatives to promote the responsible development and deployment of artificial intelligence. Instruments such as the OECD AI Principles, the UNESCO Recommendation on the Ethics of Artificial Intelligence, and the European Union Artificial Intelligence Act establish common principles including transparency, accountability, fairness, human oversight, and non-discrimination as essential requirements for trustworthy AI. These developments demonstrate an emerging international consensus that AI governance cannot be separated from the protection of human rights and democratic values. Nevertheless, existing regulatory approaches remain predominantly risk-based and compliance-oriented, concentrating on technical standards, ethical guidelines, and institutional governance while paying comparatively limited attention to the constitutional implications of algorithmic decision-making.[11]

This limitation becomes increasingly significant as AI systems assume functions that directly influence the exercise of public power and the enjoyment of constitutional rights. Although contemporary AI regulations provide important mechanisms for mitigating technological risks, they do not fundamentally address how constitutional principles should operate when decisions affecting equality, due process, and human dignity are increasingly mediated by algorithmic systems. In other words, the challenge extends beyond regulating technology itself to determining how constitutional safeguards should evolve in response to new forms of digital and algorithmic power. This constitutional dimension has encouraged the emergence of digital constitutionalism, a developing body of scholarship that seeks to preserve constitutional values within digitally mediated systems of governance. Rather than replacing traditional constitutional law, digital constitutionalism reconceptualises constitutional principles so that they remain effective in limiting both public and private forms of power exercised through digital technologies.

Against this background, gender bias in artificial intelligence should be understood not merely as a technical failure or an issue of algorithmic fairness, but as a constitutional problem requiring a broader normative response. While existing studies have extensively examined gender bias from the perspectives of computer science, AI ethics, and regulatory governance, relatively limited attention has been devoted to analysing how algorithmic discrimination affects constitutional rights within the framework of digital constitutionalism. This gap is particularly significant given the expanding role of AI in sectors that directly influence individuals' legal status, access to public services, and equal participation in society. Accordingly, this article examines how gender bias in artificial intelligence challenges constitutional rights and explores how digital constitutionalism may contribute to reconstructing constitutional safeguards capable of ensuring equality, non-discrimination, accountability, and the effective protection of fundamental rights in the age of algorithmic governance.

Existing scholarship has extensively examined artificial intelligence from multiple disciplinary perspectives, however, the literature remains fragmented in

[11] Giovanni De Gregorio, *Digital Constitutionalism in Europe: Reframing Rights and Powers in the Algorithmic Society*, 1st ed. (Cambridge University Press, 2022), https://doi.org/10.1017/9781009071215.

explaining the constitutional implications of gender-biased algorithmic decision-making. Early legal scholarship by Barocas and Selbst established that algorithmic systems may reproduce historical discrimination through *disparate impact*, demonstrating that data-driven decision-making is not inherently neutral but capable of reinforcing existing social inequalities.[12] Building upon this foundation, Feldman et al. and Friedler et al. developed computational approaches to detecting and mitigating algorithmic bias, while subsequent studies further refined algorithmic fairness through statistical optimisation and fairness-aware machine learning models.[13] These contributions significantly advanced technical solutions for reducing discriminatory outcomes but largely conceptualised fairness as a computational problem rather than a constitutional one. A parallel body of literature has specifically examined gender bias in AI. Buolamwini and Gebru, through their influential *Gender Shades* study, demonstrated substantial disparities in the performance of commercial facial recognition systems across gender and skin colour,[14] findings that were reinforced by Criado Perez, who illustrated how gender data gaps systematically disadvantage women in technology and public policy,[15] and by Heinrichs, who revealed how automated decision-making may reproduce structural inequalities within public services.[16] Collectively, these studies provide compelling empirical evidence of gender-biased algorithmic outcomes but primarily interpret the issue through technical, sociological, or ethical perspectives rather than constitutional analysis. At the governance level, Floridi et al. introduced ethical principles for trustworthy AI,[17] while Karen Yeung conceptualised algorithmic regulation as an emerging model of governance,[18] developments later reflected in international instruments such as the OECD AI Principles and the European Union Artificial Intelligence Act, both of which emphasise transparency, accountability, fairness, and human oversight.[19] Nevertheless, these regulatory frameworks remain predominantly oriented towards risk management and institutional compliance without fully addressing the constitutional transformation generated by algorithmic power. More recently, constitutional scholars including Edoardo Celeste and Oreste Pollicino have developed the concept of digital constitutionalism, arguing that constitutional values

[12] Solon Barocas and Andrew D. Selbst, "Big Data's Disparate Impact," *SSRN Electronic Journal*, ahead of print, 2016, https://doi.org/10.2139/ssrn.2477899.
[13] Michael Feldman et al., "Certifying and Removing Disparate Impact," arXiv:1412.3756, preprint, arXiv, July 16, 2015, https://doi.org/10.48550/arXiv.1412.3756.
[14] Joy Buolamwini and Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, n.d.
[15] "Invisible Women | Caroline Criado Perez," accessed July 22, 2026, https://carolinecriadoperez.com/book/invisible-women/.
[16] Bert Heinrichs, "Discrimination in the Age of Artificial Intelligence," *AI & SOCIETY* 37, no. 1 (2022): 143–54, https://doi.org/10.1007/s00146-021-01192-2.
[17] Luciano Floridi et al., "AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations," *Minds and Machines* 28, no. 4 (2018): 689–707, https://doi.org/10.1007/s11023-018-9482-5.
[18] Diver, "Karen Yeung and Martin Lodge (Eds) Algorithmic Regulation Reviewed by Laurence Diver."
[19] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying down Harmonised Rules on Artificial Intelligence and Amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA Relevance) (2024), http://data.europa.eu/eli/reg/2024/1689/oj.

must evolve to constrain new forms of digital power exercised by both public and private actors.[20] While this scholarship substantially enriches contemporary constitutional theory, its primary focus remains on digital platforms, Internet governance, and the broader constitutional architecture of the digital environment, leaving comparatively limited analysis of how digital constitutionalism may be employed to address gender-biased algorithmic decision-making. Taken together, the existing literature demonstrates significant progress in understanding algorithmic bias, AI governance, and digital constitutionalism; nevertheless, these strands of scholarship continue to develop largely in parallel, with limited integration between gender bias, constitutional rights, and digital constitutionalism as a unified analytical framework.

The foregoing literature reveals three principal limitations. First, existing studies on algorithmic bias and AI fairness predominantly conceptualise discrimination as a technical or computational problem, emphasising statistical optimisation and bias mitigation while paying comparatively limited attention to constitutional protection. Second, scholarship on gender bias successfully demonstrates the existence of discriminatory algorithmic outcomes across multiple sectors but generally approaches the issue from empirical, ethical, or human rights perspectives rather than analysing its implications for constitutional principles such as equality, due process, and human dignity. Third, although digital constitutionalism has emerged as an influential framework for understanding the constitutional governance of digital technologies, its application to gender-biased algorithmic decision-making remains insufficiently developed. As a result, the intersection between gender bias, constitutional rights, and digital constitutionalism continues to represent an important theoretical gap within contemporary legal scholarship.

Building upon this gap, this article examines how gender bias in artificial intelligence affects constitutional rights and explores how digital constitutionalism can serve as a normative framework for reconstructing constitutional safeguards against algorithmically mediated discrimination. Unlike previous studies that primarily emphasise algorithmic fairness, AI ethics, or regulatory compliance, this research integrates constitutional law, equality jurisprudence, and digital constitutionalism into a single analytical framework. By positioning gender bias as a constitutional problem arising from the exercise of algorithmic power, the article contributes to the emerging discourse on constitutional governance in the age of artificial intelligence and offers a conceptual foundation for strengthening constitutional protection in increasingly digitalised systems of public and private decision-making.

Research Methods

This study employs doctrinal legal research to examine the constitutional implications of gender bias in artificial intelligence and to develop a normative framework through digital constitutionalism. The research adopts a statutory approach by analysing international and regional AI governance instruments, including the European Union Artificial Intelligence Act, the UNESCO

[20] Edoardo Celeste, *Digital Constitutionalism: The Role of Internet Bills of Rights*, 1st ed. (Routledge, 2022), https://doi.org/10.4324/9781003256908.

Recommendation on the Ethics of Artificial Intelligence, and the OECD AI Principles, alongside relevant human rights and constitutional norms. It also employs a conceptual approach to examine the theories of algorithmic governance, algorithmic bias, constitutional rights, and digital constitutionalism. The study relies on primary and secondary legal materials, including legislation, international instruments, scholarly books, and peer-reviewed journal articles. These materials are analysed qualitatively using descriptive and prescriptive legal reasoning to evaluate existing constitutional protections and formulate a normative framework for addressing gender bias in artificial intelligence.

Gender Bias in Artificial Intelligence and Its Constitutional Implications

The rapid integration of artificial intelligence into both public governance and private decision-making has fundamentally transformed the manner in which legally significant decisions are produced. Decisions concerning employment, access to credit, healthcare, education, social welfare, and even criminal justice are increasingly mediated by algorithmic systems that operate with limited human intervention. This transformation extends beyond the digitalisation of administrative processes, reflecting a structural redistribution of decision-making authority from human institutions to algorithmic systems that increasingly mediate legally significant decisions affecting the exercise of individual rights. Consequently, computational systems no longer function merely as administrative tools but have become institutional actors capable of shaping access to rights, opportunities, and public services at an unprecedented scale.[21] Consequently, the legal significance of artificial intelligence can no longer be assessed solely in terms of technological efficiency or regulatory compliance. As algorithmic systems increasingly determine access to opportunities, public services, and legal entitlements, they also acquire the capacity to shape the practical enjoyment of constitutionally protected rights. The central constitutional concern, therefore, lies not in the existence of artificial intelligence itself but in whether the exercise of algorithmic power remains compatible with the principles of equality, accountability, transparency, and the rule of law.

Within this broader constitutional context, gender bias represents one of the most significant manifestations of algorithmic discrimination.[22] Unlike traditional forms of discrimination that generally originate from identifiable human actors, gender bias in AI frequently emerges through complex interactions between historical datasets, model design, optimisation objectives, and institutional practices embedded within algorithmic systems.[23] This characteristic makes algorithmic discrimination considerably more difficult to identify, explain, and remedy through conventional legal mechanisms. More importantly, because algorithmic decisions increasingly affect areas that directly implicate constitutional rights, gender bias should not be understood merely as a technical defect or ethical concern. Rather, it constitutes a

[21] John Danaher et al., "Algorithmic Governance: Developing a Research Agenda through the Power of Collective Intelligence," *Big Data & Society* 4, no. 2 (2017): 205395171772655, https://doi.org/10.1177/2053951717726554.
[22] "Invisible Women | Caroline Criado Perez."
[23] Sinead O'Connor and Helen Liu, "Gender Bias Perpetuation and Mitigation in AI Technologies: Challenges and Opportunities," *AI & SOCIETY* 39, no. 4 (2024): 2045–57, https://doi.org/10.1007/s00146-023-01675-4.

constitutional problem arising from the interaction between digital technologies and the unequal distribution of power within contemporary society. Understanding the nature and sources of gender bias in artificial intelligence therefore becomes an essential starting point for evaluating whether existing constitutional safeguards remain adequate to protect fundamental rights in the age of algorithmic governance.

The Nature and Sources of Gender Bias in AI

Gender bias in artificial intelligence should not be understood as an isolated technical defect but as a structural manifestation of discrimination embedded within algorithmic decision-making.[24] The widespread assumption that computational systems produce objective and neutral outcomes has increasingly been challenged by empirical evidence demonstrating that algorithms frequently inherit, reproduce, and even amplify historical inequalities contained in the data from which they learn. As Barocas and Selbst argue, algorithmic systems do not operate independently of society; rather, they derive their predictive capacity from historical data that often reflect discriminatory social practices.[25] Consequently, AI may generate disparate impacts even in the absence of any explicit intention to discriminate, making algorithmic discrimination considerably more difficult to detect and challenge than conventional forms of human decision-making.[26] The apparent neutrality of algorithms therefore conceals what is essentially a continuation of existing power asymmetries in digital form, transforming historical prejudice into automated decisions that appear scientifically objective.[27]

The existence of gender bias in AI is no longer merely theoretical. One of the most widely cited examples is Amazon's AI recruitment system, which the company abandoned after internal evaluations revealed that the algorithm systematically downgraded applications containing words associated with women, including references to women's organisations and female-only educational achievements.[28] The system had been trained using historical recruitment data dominated by male applicants over a ten-year period. Rather than identifying the most qualified candidates, the algorithm learned that male characteristics were statistically associated with successful recruitment and consequently reproduced that historical preference in future hiring recommendations.[29] The Amazon case demonstrates that AI does not independently create discrimination, instead, it legitimises historical discrimination by converting past institutional practices into automated decision-making rules.

---

[24] Rena Schwarting and Lena Ulbricht, "Why Organization Matters in 'Algorithmic Discrimination,'" *KZfSS Kölner Zeitschrift Für Soziologie Und Sozialpsychologie* 74, no. S1 (2022): 307–30, https://doi.org/10.1007/s11577-022-00838-3.

[25] Selbst et al., "Fairness and Abstraction in Sociotechnical Systems."

[26] Pauline T. Kim, *DATA-DRIVEN DISCRIMINATION AT WORK*, n.d.

[27] Hao Wang, "Transparency as Manipulation? Uncovering the Disciplinary Power of Algorithmic Transparency," *Philosophy & Technology* 35, no. 3 (2022): 69, https://doi.org/10.1007/s13347-022-00564-w.

[28] Manish Raghavan, *The Societal Impacts of Algorithmic Decision-Making*, 1st ed. (ACM, 2023), https://doi.org/10.1145/3603195.

[29] Alina Köchling and Marius Claus Wehner, "Discriminated by an Algorithm: A Systematic Review of Discrimination and Fairness by Algorithmic Decision-Making in the Context of HR Recruitment and HR Development," *Business Research* 13, no. 3 (2020): 795–848, https://doi.org/10.1007/s40685-020-00134-w.

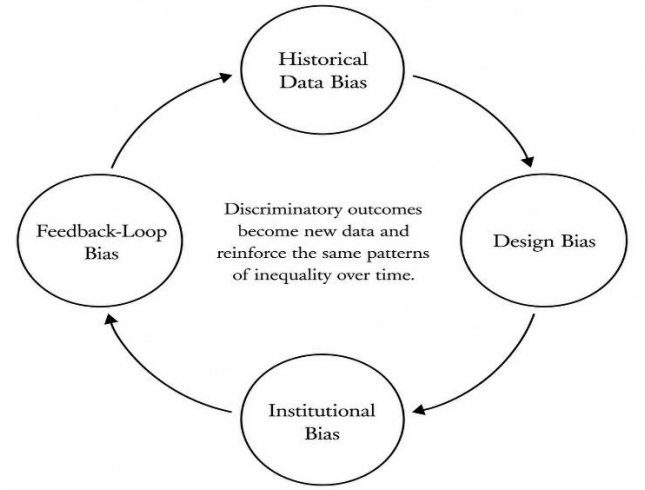
Empirical evidence supporting this phenomenon was further established through the landmark Gender Shades study conducted by Joy Buolamwini and Timnit Gebru.[30] Evaluating commercial facial recognition systems developed by major technology companies, the study found substantial disparities in gender classification accuracy across demographic groups.[31] While the systems achieved near-perfect accuracy for lighter-skinned men, error rates increased dramatically for darker-skinned women, exceeding thirty percent in several commercial products. These findings demonstrate that algorithmic performance is not determined solely by computational sophistication but is significantly influenced by the representativeness of training datasets and the assumptions underlying model development. In other words, algorithmic fairness depends not merely on technological capability but on whether the social diversity of the affected population is adequately reflected throughout the design process.

Judicial developments have likewise begun to acknowledge the constitutional implications of algorithmic decision-making. In State v. Loomis, the Wisconsin Supreme Court considered the use of the proprietary COMPAS algorithm during criminal sentencing.[32] Although the court ultimately upheld its use, it expressly recognised that algorithmic risk assessment presents significant concerns regarding transparency, explainability, and procedural fairness because defendants are unable to examine the underlying methodology used to generate risk scores. The case illustrates that algorithmic systems increasingly exercise legal influence over decisions affecting individual rights while remaining largely insulated from meaningful judicial scrutiny due to claims of commercial confidentiality. Although *Loomis* concerned criminal sentencing rather than gender discrimination, its constitutional significance extends beyond the facts of the case by demonstrating how algorithmic decision-making may undermine due process and accountability when individuals cannot effectively challenge automated reasoning.

From these developments, it becomes evident that gender bias originates from multiple and interconnected sources rather than from data alone. *First*, historical data bias arises because machine-learning models learn from datasets reflecting longstanding patterns of gender inequality in employment, finance, healthcare, education, and criminal justice. When historical decisions themselves are discriminatory, algorithms inevitably reproduce those patterns unless corrective interventions are deliberately introduced. *Second*, design bias emerges through choices made by developers regarding variable selection, optimisation objectives, model architecture, and performance metrics. Systems optimised exclusively for predictive accuracy may inadvertently privilege statistical efficiency over substantive equality.[33] *Third,* institutional bias reflects the governance environment within which AI systems are developed and deployed. Algorithmic models frequently encode organisational priorities, economic incentives, and regulatory assumptions that mirror existing institutional inequalities. Finally, feedback-loop bias enables discriminatory outcomes to perpetuate themselves over time. Decisions generated by

[30] Buolamwini and Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*.
[31] Alice J. O'Toole et al., *Demographic Effects on Estimates of Automatic Face Recognition Performance*, n.d.
[32] *State v. Loomis*, **2016 WI 68**, 371 Wis. 2d 235, 881 N.W.2d 749 (Wis. 2016).
[33] Mitchell et al., "Algorithmic Fairness."

algorithms subsequently become new training data, reinforcing the very patterns of inequality upon which future predictions are based. This recursive process transforms discrimination from an isolated event into a self-reinforcing system of automated exclusion.[34]



These characteristics distinguish gender bias in AI from conventional discrimination in an important constitutional respect. Traditional constitutional doctrine generally assumes that discriminatory treatment can be attributed to identifiable state actors or private individuals whose conduct is capable of judicial review. Algorithmic discrimination, by contrast, emerges through complex interactions among data providers, software developers, technology companies, and public institutions deploying AI-assisted systems.[35] Responsibility consequently becomes fragmented across multiple actors, while discriminatory outcomes appear as the product of technical objectivity rather than human choice. This diffusion of accountability creates a constitutional gap in which infringements of equality and non-discrimination may persist without an easily identifiable legal subject responsible for the violation. Accordingly, the principal challenge posed by gender-biased AI lies not merely in correcting inaccurate algorithms after harm occurs but in ensuring that the design, governance, and deployment of algorithmic systems remain subject to constitutional principles of equality, transparency, accountability, and the rule of law.[36]

Constitutional Right Affected by Gender-Biased AI

The constitutional implications of gender-biased artificial intelligence extend far beyond questions of technological accuracy or regulatory compliance. Once algorithmic systems are entrusted with making or supporting decisions concerning employment, education, healthcare, access to financial services, law enforcement, and judicial administration, discriminatory outputs cease to be merely technical

[34] Lydia T. Liu et al., "Delayed Impact of Fair Machine Learning," *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, August 2019, 6196–200, https://doi.org/10.24963/ijcai.2019/862.
[35] Köchling and Wehner, "Discriminated by an Algorithm."
[36] Danielle Keats Citron and Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, n.d.

errors and become potential infringements of constitutionally protected rights. Unlike traditional discrimination, which generally involves identifiable human decision-makers whose actions can be scrutinised through conventional legal procedures, algorithmic discrimination often operates through opaque computational processes whose reasoning remains inaccessible to affected individuals. Consequently, gender bias in AI challenges not only the substantive guarantee of equality but also the constitutional architecture through which individuals exercise, protect, and enforce their fundamental rights. This transformation requires constitutional law to confront a new form of public and private power exercised through automated decision-making rather than through conventional governmental action.

The first constitutional right directly affected is the right to equality and non-discrimination, universally recognised as one of the foundational principles of constitutional democracies. Constitutional equality requires that individuals in comparable situations be treated equally unless differential treatment can be objectively and reasonably justified.[37] Gender-biased AI undermines this principle by producing systematically different outcomes for women despite the apparent neutrality of algorithmic processes. The constitutional concern therefore lies not merely in explicit discrimination but in indirect or structural discrimination, where neutral computational models disproportionately disadvantage one group because they reproduce historical patterns embedded within training data.[38] The Amazon recruitment system illustrates this phenomenon, although the algorithm never received explicit instructions to discriminate against women, it learned from historically male-dominated recruitment records and consequently downgraded female applicants.[39] Constitutionally, such outcomes challenge the principle of equal protection because discrimination arises not from intentional state action but from algorithmic optimisation that perpetuates existing social inequalities. This evolution demonstrates that constitutional equality must increasingly address algorithmic disparate impact, rather than limiting its focus to intentional discrimination alone.

Gender-biased AI also threatens the constitutional protection of human dignity, which increasingly functions as the normative foundation of digital constitutionalism. Human dignity presupposes that every individual possesses intrinsic worth and should never be reduced to an object of automated calculation. Yet algorithmic systems frequently transform individuals into statistical profiles evaluated according to probabilistic predictions rather than their personal circumstances.[40] Women affected by biased hiring algorithms, facial recognition technologies, or automated credit-scoring systems are therefore not merely subjected to inaccurate decisions, they are deprived of meaningful recognition as autonomous rights-bearing individuals. As De Gregorio and Celeste argue, the constitutional message underlying the GDPR is fundamentally human-centred, requiring technological innovation to remain subordinate to human autonomy and dignity rather than economic efficiency or computational optimisation.[41] From this perspective, gender bias constitutes a constitutional injury because it objectifies

[37] Elise Muir, "The Essence of the Fundamental Right to Equal Treatment: Back to the Origins," *German Law Journal* 20, no. 6 (2019): 817–39, https://doi.org/10.1017/glj.2019.64.
[38] Barocas and Selbst, "Big Data's Disparate Impact."
[39] Raghavan, *The Societal Impacts of Algorithmic Decision-Making*, 1st ed. (ACM, 2023).
[40] Citron and Pasquale, *The Scored Society: Due Process for Automated Predictions*.
[41] Celeste, *Digital Constitutionalism*, 1st ed.

individuals through data-driven classifications that reinforce pre-existing stereotypes instead of respecting their equal moral and legal status.[42]

A further constitutional concern arises in relation to due process and procedural fairness. Constitutional guarantees of due process traditionally require that individuals understand the reasons underlying decisions affecting their rights and possess effective opportunities to challenge those decisions before an impartial authority. These guarantees become increasingly fragile when decisions are generated through opaque algorithmic models. The decision of the Wisconsin Supreme Court in State v. Loomis illustrates this dilemma. Although the court permitted the use of the COMPAS risk-assessment algorithm during sentencing, it acknowledged that defendants were unable to examine or meaningfully challenge the proprietary methodology used to calculate risk scores. Although the dispute did not arise from gender discrimination, the judgment recognised that algorithmic decision-making poses constitutional concerns because procedural fairness cannot be fully realised where individuals are unable to understand, verify, or challenge the logic embedded within computational systems. The constitutional significance of *Loomis* therefore extends beyond criminal justice by demonstrating that algorithmic opacity itself may compromise fundamental guarantees of due process wherever AI influences legally significant decisions.

Gender-biased AI affects the constitutional principle of access to an effective remedy and accountability. Constitutional rights possess little practical value unless individuals can identify the actor responsible for the infringement and obtain meaningful legal redress. Algorithmic decision-making complicates this requirement because responsibility is frequently dispersed among software developers, data providers, technology companies, and public institutions deploying AI systems. As a consequence, discriminatory outcomes often occur without any single actor accepting legal responsibility. This diffusion of accountability creates a constitutional deficit in which violations of equality, dignity, and due process may persist despite the absence of an effective mechanism for judicial review.[43] Recent scholarship therefore argues that constitutional protection in the algorithmic society should no longer focus exclusively on limiting state power but must also establish constitutional constraints upon private digital actors exercising quasi-public decision-making authority.[44] In this respect, gender-biased AI exposes a broader transformation in constitutional governance, the principal threat to fundamental rights increasingly originates not only from governmental institutions but also from algorithmic infrastructures developed and controlled by private technology companies.

These developments demonstrate that gender bias in AI is not simply a question of technological fairness but a constitutional problem affecting the core values upon which modern constitutional democracies are founded. Equality, human dignity, procedural fairness, and effective judicial protection are no longer threatened solely by discriminatory legislation or arbitrary governmental action, but also by opaque algorithmic systems that increasingly mediate access to rights and opportunities. Accordingly, the constitutional challenge is not merely to prohibit discriminatory outcomes after they occur but to ensure that the design, deployment,

[42] De Gregorio and Radu, "Digital Constitutionalism in the New Era of Internet Governance."
[43] Deborah G. Johnson, "ALGORITHMIC ACCOUNTABILITY *IN THE MAKING*," *Social Philosophy and Policy* 38, no. 2 (2021): 111–27, https://doi.org/10.1017/S0265052522000073.
[44] De Gregorio and Radu, "Digital Constitutionalism in the New Era of Internet Governance."

and governance of artificial intelligence remain subject to constitutional principles capable of constraining algorithmic power in the digital age. This constitutional imperative provides the foundation for evaluating whether existing AI governance frameworks are sufficient to address gender-biased algorithmic decision-making.

Reconstructing Constitutional Protection Through Digital Constitutionalism

Gender-biased artificial intelligence is not simply a technological flaw but a constitutional challenge resulting from the growing exercise of algorithmic power over fundamental rights. As AI systems increasingly influence decisions concerning equality, human dignity, procedural fairness, and access to effective remedies, constitutional law can no longer rely solely on its traditional role of reviewing unlawful actions after harm has occurred. Instead, constitutional protection must adapt to a governance environment in which decisions with significant legal consequences are increasingly produced through algorithmic systems operating across both public institutions and private digital platforms. The key constitutional issue is therefore not only whether algorithmic discrimination exists, but whether existing constitutional frameworks remain capable of limiting algorithmic power before violations of fundamental rights occur.

Addressing this challenge requires moving beyond merely identifying constitutional violations towards reconstructing constitutional safeguards that are appropriate for the digital age. Although existing AI governance frameworks have introduced important principles such as transparency, accountability, fairness, and human oversight, they continue to regulate AI primarily as a technological risk rather than as a constitutional issue.[45] Consequently, these frameworks remain insufficient to address the structural redistribution of power generated by algorithmic decision-making. A constitutional approach is therefore needed to ensure that the development and deployment of AI remain consistent with constitutional principles and provide effective protection for fundamental rights. Against this background, the following discussion examines the limitations of existing AI governance and constitutional frameworks and proposes a constitutional model for protecting fundamental rights against gender-biased artificial intelligence.

Limitations of Existing AI Governance and Constitutional Frameworks

The adoption of artificial intelligence in Indonesia has progressed more rapidly than the development of its legal framework. According to the *Work Trend Index 2024*, 92% of business leaders in Indonesia consider AI adoption essential for maintaining competitiveness, exceeding both the global average of 79% and the Asia-Pacific average of 84%.[46] However, 48% of respondents also reported that their organisations had not yet established a clear vision or adequate implementation strategy for AI adoption. By 2025, the use of AI had expanded across the workforce. The *PwC Global Workforce Hopes and Fears Survey 2025*, published in February 2026, found that 69% of Indonesian workers had used AI in their jobs during the preceding

[45] De Gregorio, *Digital Constitutionalism in Europe*.
[46] Microsoft and LinkedIn, *Work Trend Index 2024: AI at Work in Indonesia* (Jakarta: Microsoft Indonesia, 11 June 2024), https://news.microsoft.com/id-id/2024/06/11/microsoft-dan-linkedin-luncurkan-work-trend-index-2024-menilik-keadaan-ai-dalam-dunia-kerja-di-indonesia/

twelve months.[47] At the same time, the Indonesian government has begun integrating AI into public services, including healthcare, administrative decision-making, and various national priority programmes.[48] These developments indicate that AI is no longer confined to enhancing corporate productivity but is increasingly becoming a critical infrastructure shaping public service delivery and the allocation of public resources. Nevertheless, as of July 2026, Indonesia had not yet enacted a dedicated legal framework comprehensively governing the development, deployment, and accountability of artificial intelligence systems. The proposed Presidential Regulation on the National AI Roadmap and AI Governance remained under preparation, leaving the existing regulatory framework fragmented across constitutional provisions, personal data protection law, electronic system regulations, and ethical guidelines that are not uniformly legally binding.

At the constitutional level, the 1945 Constitution of the Republic of Indonesia provides a strong normative foundation for addressing algorithmic discrimination. Article 27(1) guarantees equality before the law and government, while Article 28D(1) guarantees the right to recognition, protection, and equal treatment before the law, as well as legal certainty founded upon justice. Article 28H(2) recognises the right to receive special measures in order to achieve equality and justice, whereas Article 28I(2) expressly prohibits discriminatory treatment on any grounds. These constitutional guarantees provide a normative basis for understanding gender discrimination generated by artificial intelligence as a constitutional issue rather than merely a technical failure. Accordingly, the principal strength of Indonesia's constitutional framework lies in its broad commitment to equality, non-discrimination, and substantive justice, enabling these principles to be applied to emerging technological developments, including artificial intelligence.

Its limitation, however, lies in a constitutional framework that remains largely structured around conventional forms of legal violations. Constitutional norms are generally designed to address conduct attributable to identifiable state institutions or legal subjects, whereas algorithmic decisions emerge from far more complex interactions among data providers, model developers, electronic system operators, user institutions, and automated systems themselves. Consequently, the Constitution does not provide operational standards for determining when the use of AI constitutes a violation of the principle of equality, how discriminatory impacts should be established, who should bear legal responsibility, or what forms of legal remedy should be available. As a result, constitutional guarantees continue to operate predominantly in an *ex post* manner, providing protection only after harm has occurred rather than imposing preventive obligations throughout the design, training, testing, and deployment of AI systems.

The legal instrument most directly relevant to AI governance in Indonesia is Law No. 27 of 2022 on Personal Data Protection (PDP Law). The PDP Law establishes the rights of data subjects and the obligations of data controllers and,

[47] PricewaterhouseCoopers (PwC), *Global Workforce Hopes and Fears Survey 2025* (Jakarta: PwC Indonesia, 23 February 2026), https://www.pwc.com/id/en/media-centre/press-release/2026/english/hopes-and-fears-2025.html

[48] Kementerian Komunikasi dan Digital Republik Indonesia, "Lima Prioritas Pemerintah untuk Pengembangan AI," 14 Januari 2025, diakses 3 Agustus 2026, https://www.komdigi.go.id/berita/masyarakat-digital/detail/lima-prioritas-pemerintah-untuk-pengembangan-ai

through Article 10, grants individuals the right to object to decisions based solely on automated processing, while Article 12 guarantees the right to seek legal remedies and compensation for unlawful data processing. These provisions demonstrate that Indonesian law has begun to recognise the legal risks associated with automated decision-making. Nevertheless, the protection afforded by the PDP Law remains limited because it applies only to decisions based exclusively on automated processing and primarily focuses on the protection of personal data. It does not establish obligations relating to gender bias testing, independent algorithmic audits, dataset representativeness, or the transparency and explainability of AI systems. Consequently, although the PDP Law provides an important foundation for AI governance, it remains insufficient to ensure effective protection against gender discrimination arising throughout the lifecycle of artificial intelligence systems.

The next relevant legal framework consists of Law No. 11 of 2008 on Electronic Information and Transactions, as most recently amended by Law No. 1 of 2024, together with Government Regulation No. 71 of 2019 on the Operation of Electronic Systems and Transactions. The principal strength of these instruments lies in establishing fundamental obligations for Electronic System Operators. Article 3 of Government Regulation No. 71 of 2019 requires every operator to ensure that electronic systems are operated in a reliable, secure, and accountable manner. The regulation further establishes obligations concerning system security, data governance, and responsibility for the proper operation of electronic systems. These provisions are particularly relevant because AI systems deployed by public institutions and private companies generally function as components of broader electronic systems.

Nevertheless, the concepts of "reliability" and "security" embodied in these regulations primarily concern the functional performance, operational security, and continuity of electronic systems rather than the fairness of their outcomes. An algorithm may therefore operate with a high degree of technical accuracy while still producing discriminatory effects against women. Neither the Electronic Information and Transactions Law nor Government Regulation No. 71 of 2019 establishes a risk-based classification of AI systems, prohibits specific high-risk AI practices, or requires fundamental rights impact assessments, algorithmic bias audits, model transparency, or human review of automated decisions. Nor do these instruments distinguish conventional electronic systems from AI systems used to determine access to employment, credit, healthcare, education, social assistance, or other public services. Consequently, the regulatory framework continues to conceive the responsibility of Electronic System Operators primarily in terms of system security and operational reliability, rather than addressing the exercise of algorithmic power that may adversely affect constitutional rights.

Indonesia has also adopted the National Strategy for Artificial Intelligence (Stranas AI) 2020–2045.[49] The Strategy identifies healthcare, bureaucratic reform, education and research, food security, and smart mobility and cities as the priority sectors for AI development. Its principal strength lies in providing a long-term policy direction, aligning technological development with national interests, and

[49] Badan Pengkajian dan Penerapan Teknologi (BPPT), *Strategi Nasional Kecerdasan Artifisial Indonesia 2020–2045* (Jakarta: BPPT, 2020), https://library.mastel.or.id/wp-content/uploads/2021/06/stranas-ka.pdf

recognising ethics as one of the core pillars of Indonesia's AI ecosystem. It also reflects the government's awareness of the need for coordinated AI governance, talent development, digital infrastructure, and research. Nevertheless, the National Strategy is a policy document rather than a legally binding regulatory instrument capable of creating enforceable rights and obligations. It neither establishes a dedicated supervisory authority nor provides procedures for algorithmic auditing, complaint mechanisms, sanctions, or legal remedies for individuals adversely affected by AI systems. Moreover, its primary focus remains on fostering innovation and enhancing national competitiveness rather than constraining AI applications that may threaten constitutional rights. Consequently, while the National Strategy serves as an important roadmap for AI development, it cannot function as a direct legal basis for reviewing or invalidating discriminatory algorithmic decisions.

A more specific initiative in Indonesia's AI governance framework emerged with the issuance of the Minister of Communication and Informatics Circular Letter No. 9 of 2023 on Artificial Intelligence Ethics, which provides ethical guidance for AI-based business actors and Electronic System Operators in both the public and private sectors. The Circular introduces principles of inclusiveness, accessibility, security, humanity, transparency, credibility, and accountability, while emphasising that the development and deployment of AI should take account of social impacts, avoid discrimination, protect privacy, and remain subject to meaningful human oversight. Compared with earlier regulatory instruments, the Circular represents a significant step forward by explicitly recognising the ethical dimensions of AI governance. Nevertheless, as a soft law instrument, its legal effect remains limited because it does not establish binding technical standards, independent auditing mechanisms, regulatory oversight procedures, or enforceable sanctions to ensure compliance. Moreover, principles such as inclusiveness, transparency, and accountability have not been translated into concrete operational obligations, including mandatory gender bias testing, the representation of women in training datasets, or the duty to provide meaningful explanations for algorithmic decisions. Consequently, the Circular functions primarily as a normative ethical guideline rather than an effective legal instrument capable of preventing gender discrimination throughout the design, development, and deployment of AI systems.

Another limitation arises from Indonesia's fragmented regulatory approach to AI governance. The 1945 Constitution guarantees the principles of equality and non-discrimination, the Personal Data Protection Law governs data processing and certain forms of automated decision-making, the Electronic Information and Transactions Law together with Government Regulation No. 71 of 2019 regulates electronic systems, while the National Strategy for Artificial Intelligence 2020–2045 and the Minister of Communication and Informatics Circular Letter No. 9 of 2023 provide policy direction and ethical guidance. Although each instrument performs an important regulatory function, none establishes a comprehensive framework that integrates constitutional principles, obligations for AI developers, the responsibilities of deploying institutions, independent oversight mechanisms, and effective remedies for individuals affected by algorithmic discrimination. As a result, an AI system may comply with personal data protection and electronic systems regulations while still producing discriminatory outcomes based on gender.

The Indonesian government's plan to enact a Presidential Regulation on AI governance and the National AI Roadmap demonstrates an effort to move beyond

voluntary guidelines towards a more legally binding regulatory framework. The proposed regulation is reported to include provisions on the deployment of AI across central and local government institutions, risk assessment mechanisms, safeguards against the misuse of biometric technologies and deepfakes, as well as the use of AI in national priority programmes. These developments represent a positive step by recognising that the expanding deployment of AI must be accompanied by appropriate risk management mechanisms. Nevertheless, as the proposed Presidential Regulation remained in draft form as of July 2026, its final content and its effectiveness in protecting fundamental rights cannot yet be assessed with certainty. Moreover, a risk-based regulatory approach alone may prove insufficient unless it expressly establishes equality, non-discrimination, human dignity, due process, and effective remedies as autonomous legal obligations governing the entire lifecycle of artificial intelligence systems.

The principal weakness of Indonesia's AI governance does not lie merely in the absence of a dedicated artificial intelligence statute, but in the lack of a coherent normative framework connecting technology regulation with constitutional protection.[50] The existing legal framework provides an important foundation through constitutional guarantees of equality, personal data protection, electronic systems security, and ethical principles for AI. However, these foundations have not been complemented by explicit prohibitions against algorithmic discrimination, mandatory constitutional or fundamental rights impact assessments, pre-deployment gender bias testing, independent algorithmic audits, meaningful transparency, human oversight, clear allocation of legal responsibility, or accessible mechanisms for effective remedies. Consequently, the current framework remains predominantly sectoral, individualised, and *ex post* in nature, whereas the risks posed by AI arise throughout the entire lifecycle of algorithmic systems, from data selection and model design to optimisation objectives, deployment, and real-world application. Accordingly, the legal challenge is not simply to enact a dedicated AI law, but to reconstruct AI governance so that the exercise of algorithmic power is subject to constitutional principles from the design stage through post-deployment evaluation.

Digital Constitutionalism as a Constitutional Response to Algorithmic Power

The rapid advancement of artificial intelligence demonstrates that gender discrimination is no longer produced solely through human decision-making but increasingly emerges from algorithmic processes operating through data, predictive models, and automated decision-making systems. This transformation has fundamentally altered the exercise of power affecting the enjoyment of constitutional rights, shifting the source of potential discrimination from identifiable human actors to opaque computational systems. From Roscoe Pound's perspective, law functions as *a tool of social engineering*, requiring the legal system not only to adapt to social change but also to actively guide social development towards justice. Consequently, the emergence of algorithmic power requires constitutional law to evolve beyond its traditional focus on state action by developing legal mechanisms capable of constraining discriminatory AI systems before they produce constitutionally harmful

[50] Yudistira Nugroho and Syafira Antari, "Rethinking the Governance of Artificial Intelligence Oversight in Indonesia from a Global Perspective: Rekonstruksi Tata Kelola Pengawasan Artificial Intelligence Di Indonesia Dalam Perspektif Global," *Mendapo: Journal of Administrative Law* 7, no. 2 (2026): 269–87, https://doi.org/10.22437/mendapo.v7i2.55745.

outcomes.[51] Accordingly, as decision-making is increasingly delegated to AI systems, the law can no longer confine its regulatory function to the actions of states or individuals, but must also constrain the exercise of algorithmic power. This need becomes particularly urgent because gender bias in AI rarely arises from explicitly discriminatory rules; rather, it emerges through the reproduction of historical inequalities embedded in training datasets, variable selection, and algorithmic optimisation processes that generate disproportionately adverse outcomes for women. Under these circumstances, constitutional protection that operates only after a violation has occurred (*ex post*) is no longer sufficient. Gender bias in AI should therefore be understood not merely as a technical defect, but as a constitutional challenge that requires the reconstruction of legal mechanisms capable of governing new forms of digital power before they result in the infringement of fundamental rights.

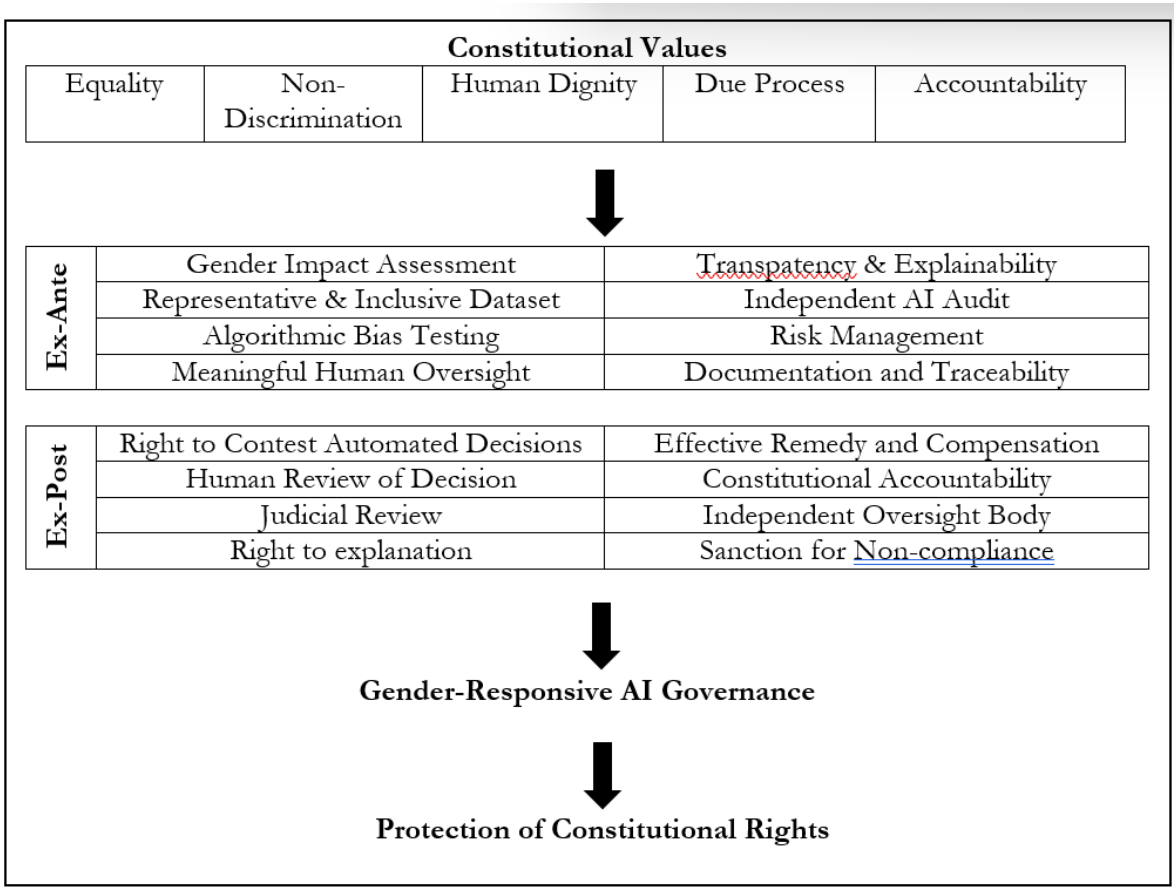
In this context, digital constitutionalism, as developed by Edoardo Celeste[52] and further elaborated by Giovanni De Gregorio[53], provides an important constitutional framework for addressing gender-biased AI. Rather than limiting constitutional constraints to digital platforms or cyberspace, this approach extends them to **algorithmic power**, recognising that AI systems increasingly shape access to employment, education, healthcare, social assistance, and other constitutionally protected rights. Accordingly, the constitutional challenge lies not in the use of AI itself, but in preventing algorithms from reproducing gender inequality without effective constitutional safeguards. Digital constitutionalism therefore requires constitutional principles—including equality, non-discrimination, transparency, accountability, meaningful human oversight, and due process—to be embedded throughout the entire AI lifecycle, from data collection and model development to algorithmic testing and deployment, thereby transforming constitutional protection from a predominantly *ex post* remedy into a preventive governance framework capable of constraining algorithmic power before discriminatory harm occurs.

A Digital Constitutionalism Model for Protecting Constitutional Rights Against Gender-Biased AI

The principal challenge posed by gender-biased AI does not lie merely in algorithmic inaccuracy, but in the absence of a constitutional framework capable of constraining algorithmic power from the moment AI systems are designed. Both Indonesia's existing legal framework and prevailing international regulatory instruments continue to treat AI primarily as an object of technology regulation, with protection largely centred on administrative compliance, system security, and risk mitigation. As a result, constitutional principles have yet to function as binding normative standards governing the entire lifecycle of AI development. This article therefore proposes a Gender-Responsive Digital Constitutionalism Model, a governance framework that places the protection of constitutional rights at the centre of the design, development, deployment, and oversight of AI systems. Unlike conventional risk-based governance, which responds only after risks have been

[51] Roscoe Pound, "The Limits of Effective Legal Action," *The International Journal of Ethics* 27, no. 2 (1917): 150–67, https://doi.org/10.1086/intejethi.27.2.2377288.
[52] Edoardo Celeste, "Digital Constitutionalism: A New Systematic Theorisation," *International Review of Law, Computers & Technology* 33, no. 1 (2019): 76–99, https://doi.org/10.1080/13600869.2019.1562604.
[53] Celeste, *Digital Constitutionalism*, 1st ed.

identified, this model proceeds from the premise that any AI system capable of affecting the exercise of constitutional rights should, from its inception, be subject to constitutional constraints.



The first layer of the proposed model is the **constitutional foundation**. All AI systems deployed in both the public and private sectors that influence access to employment, education, healthcare, credit, social assistance, and public services must be governed by the constitutional principles of equality before the law, non-discrimination, human dignity, due process, transparency, and accountability. These principles should not be treated as voluntary ethical aspirations, but as binding legal obligations applicable to all actors within the AI ecosystem. Accordingly, AI developers, data providers, technology companies, and deploying institutions bear a shared constitutional responsibility to ensure that AI systems do not produce or perpetuate gender-based discrimination.

The second layer is constitutional-by-design governance. Unlike conventional approaches that assess only the outcomes of algorithmic decision-making, this model requires constitutional safeguards to be embedded throughout the entire AI lifecycle, including data collection, model development, testing, deployment, and post-deployment evaluation. High-risk AI systems should be subject to mandatory gender impact assessments, periodic bias testing, independent algorithmic audits, comprehensive dataset documentation, and verification of the adequate representation of women in training datasets. In addition, any AI-assisted decision affecting constitutional rights must remain subject to meaningful human oversight and be accompanied by a right to explanation for affected individuals. Consequently, the model shifts constitutional protection from a reactive mechanism of redress to a

preventive governance framework in which gender discrimination is anticipated and mitigated from the earliest stages of AI system design.

The third layer is constitutional accountability. One of the principal weaknesses of AI governance lies in the fragmentation of responsibility among AI developers, data providers, technology companies, and deploying institutions. The proposed model therefore adopts the principle of shared constitutional responsibility, under which constitutional obligations are allocated across all actors involved throughout the AI lifecycle. Each actor must be able to demonstrate that the AI systems they develop, deploy, or operate comply with the constitutional standards of equality, non-discrimination, and the protection of fundamental rights. Furthermore, individuals adversely affected by algorithmic decisions should be guaranteed access to effective complaint mechanisms, meaningful human review, appropriate compensation, and oversight by an independent supervisory authority. Through this approach, constitutional protection no longer depends solely on establishing the fault of a single actor, but is instead secured through an integrated accountability framework capable of addressing the distributed nature of algorithmic power.

The proposed model demonstrates that digital constitutionalism is not merely concerned with extending constitutional principles into the digital sphere, but with fundamentally reconfiguring how constitutional law constrains power in the age of artificial intelligence. Its principal contribution lies in shifting constitutional protection from a predominantly *ex post* paradigm towards a **constitutional-by-design** approach, under which constitutional principles become binding normative standards that must be satisfied before AI systems are deployed. Accordingly, the protection of women no longer depends solely on post hoc legal remedies after discriminatory harm has occurred, but is secured through an AI governance framework deliberately designed to prevent the reproduction of gender bias from the outset. More fundamentally, the model illustrates that effective constitutional protection against gender-biased AI cannot be achieved through technology regulation alone, but requires the constitutionalisation of the entire AI lifecycle, ensuring that the design, development, deployment, and governance of algorithmic systems remain subject to constitutional constraints as a new form of power in the digital society.

The proposed model demonstrates that constitutional protection against gender-biased AI cannot be secured through a single legal instrument or a single stage of regulatory oversight, but requires the integration of constitutional safeguards throughout the entire AI lifecycle. Within this framework, constitutional principles no longer serve merely as standards for reviewing the legality of algorithmic decisions after fundamental rights have been infringed; instead, they function as binding normative standards governing the design, development, deployment, and evaluation of AI systems. Such an approach enables a more gender-responsive model of AI governance by integrating constitutional rights protection with preventive bias mitigation, continuous oversight, and the shared accountability of all actors involved in the development and deployment of AI. Accordingly, the proposed reconstruction not only extends the scope of digital constitutionalism but also provides a conceptual framework for AI governance that places the protection of women's constitutional rights at the centre of algorithmic decision-making.

Conclusion

Artificial intelligence has transformed gender discrimination from an individual and identifiable act into a systemic form of algorithmic decision-making that directly affects the exercise of constitutional rights. This study demonstrates that gender-biased AI constitutes not merely a technical or ethical problem, but a constitutional challenge arising from the growing exercise of algorithmic power over equality, human dignity, due process, and effective legal remedies. Although Indonesia has established several regulatory instruments related to personal data protection, electronic systems, and AI ethics, the existing framework remains fragmented and predominantly oriented towards technological governance rather than constitutional protection. Consequently, current regulation has not yet provided preventive safeguards capable of addressing gender discrimination throughout the AI lifecycle. Building upon the framework of digital constitutionalism, this study proposes the Gender-Responsive Digital Constitutionalism Model, which reconstructs constitutional protection by integrating constitutional values into every stage of AI development, deployment, and oversight through constitutional-by-design, gender-responsive safeguards, and constitutional accountability. This model shifts constitutional protection from an ex post mechanism that responds after rights violations occur towards an ex ante framework that prevents discriminatory algorithmic outcomes before they materialise. Accordingly, the principal contribution of this study lies in repositioning AI governance from a technology-centred regulatory approach to a constitutional governance framework that places the limitation of algorithmic power and the protection of women's constitutional rights at the centre of AI regulation.

Bibliography

Barocas, Solon, and Andrew D. Selbst. "Big Data's Disparate Impact." *SSRN Electronic Journal*, ahead of print, 2016. https://doi.org/10.2139/ssrn.2477899.

Buolamwini, Joy, and Timnit Gebru. *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*. n.d.

Burrell, Jenna. "How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms." *Big Data & Society* 3, no. 1 (2016): 2053951715622512. https://doi.org/10.1177/2053951715622512.

Caliskan, Aylin, Joanna J. Bryson, and Arvind Narayanan. "Semantics Derived Automatically from Language Corpora Contain Human-like Biases." *Science* 356, no. 6334 (2017): 183–86. https://doi.org/10.1126/science.aal4230.

Celeste, Edoardo. "Digital Constitutionalism: A New Systematic Theorisation." *International Review of Law, Computers & Technology* 33, no. 1 (2019): 76–99. https://doi.org/10.1080/13600869.2019.1562604.

Celeste, Edoardo. *Digital Constitutionalism: The Role of Internet Bills of Rights*. 1st ed. Routledge, 2022. https://doi.org/10.4324/9781003256908.

Citron, Danielle Keats, and Frank Pasquale. *The Scored Society: Due Process for Automated Predictions*. n.d.

Danaher, John, Michael J. Hogan, Chris Noone, et al. "Algorithmic Governance: Developing a Research Agenda through the Power of Collective Intelligence." *Big Data & Society* 4, no. 2 (2017): 205395171772655. https://doi.org/10.1177/2053951717726554.

De Gregorio, Giovanni. *Digital Constitutionalism in Europe: Reframing Rights and Powers in the Algorithmic Society*. 1st ed. Cambridge University Press, 2022. https://doi.org/10.1017/9781009071215.

De Gregorio, Giovanni, and Roxana Radu. "Digital Constitutionalism in the New Era of Internet Governance." *International Journal of Law and Information Technology* 30, no. 1 (2022): 68–87. https://doi.org/10.1093/ijlit/eaac004.

Diver, Laurence. "Karen Yeung and Martin Lodge (Eds) Algorithmic Regulation Reviewed by Laurence Diver." *Prometheus* 37, no. 4 (2021). https://doi.org/10.13169/prometheus.37.4.0387.

Feldman, Michael, Sorelle Friedler, John Moeller, Carlos Scheidegger, and Suresh Venkatasubramanian. "Certifying and Removing Disparate Impact." arXiv:1412.3756. Preprint, arXiv, July 16, 2015. https://doi.org/10.48550/arXiv.1412.3756.

Ferrari, Fabian, and Mark Graham. "Fissures in Algorithmic Power: Platforms, Code, and Contestation." *Cultural Studies* 35, nos. 4–5 (2021): 814–32. https://doi.org/10.1080/09502386.2021.1895250.

Floridi, Luciano, Josh Cowls, Monica Beltrametti, et al. "AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations." *Minds and Machines* 28, no. 4 (2018): 689–707. https://doi.org/10.1007/s11023-018-9482-5.

Heinrichs, Bert. "Discrimination in the Age of Artificial Intelligence." *AI & SOCIETY* 37, no. 1 (2022): 143–54. https://doi.org/10.1007/s00146-021-01192-2.

"Invisible Women | Caroline Criado Perez." Accessed July 22, 2026. https://carolinecriadoperez.com/book/invisible-women/.

Johnson, Deborah G. "ALGORITHMIC ACCOUNTABILITY *IN THE MAKING*." *Social Philosophy and Policy* 38, no. 2 (2021): 111–27. https://doi.org/10.1017/S0265052522000073.

Kim, Pauline T. *DATA-DRIVEN DISCRIMINATION AT WORK*. n.d.

Köchling, Alina, and Marius Claus Wehner. "Discriminated by an Algorithm: A Systematic Review of Discrimination and Fairness by Algorithmic Decision-

Making in the Context of HR Recruitment and HR Development." *Business Research* 13, no. 3 (2020): 795–848. https://doi.org/10.1007/s40685-020-00134-w.

Liu, Lydia T., Sarah Dean, Esther Rolf, Max Simchowitz, and Moritz Hardt. "Delayed Impact of Fair Machine Learning." *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, August 2019, 6196–200. https://doi.org/10.24963/ijcai.2019/862.

Mitchell, Shira, Eric Potash, Solon Barocas, Alexander D'Amour, and Kristian Lum. "Algorithmic Fairness: Choices, Assumptions, and Definitions." *Annual Review of Statistics and Its Application* 8, no. 1 (2021): 141–63. https://doi.org/10.1146/annurev-statistics-042720-125902.

Muir, Elise. "The Essence of the Fundamental Right to Equal Treatment: Back to the Origins." *German Law Journal* 20, no. 6 (2019): 817–39. https://doi.org/10.1017/glj.2019.64.

Nugroho, Yudistira, and Syafira Antari. "Rethinking the Governance of Artificial Intelligence Oversight in Indonesia from a Global Perspective: Rekonstruksi Tata Kelola Pengawasan Artificial Intelligence Di Indonesia Dalam Perspektif Global." *Mendapo: Journal of Administrative Law* 7, no. 2 (2026): 269–87. https://doi.org/10.22437/mendapo.v7i2.55745.

Obermeyer, Ziad, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations." *Science* 366, no. 6464 (2019): 447–53. https://doi.org/10.1126/science.aax2342.

O'Connor, Sinead, and Helen Liu. "Gender Bias Perpetuation and Mitigation in AI Technologies: Challenges and Opportunities." *AI & SOCIETY* 39, no. 4 (2024): 2045–57. https://doi.org/10.1007/s00146-023-01675-4.

O'Toole, Alice J., P. Jonathon Phillips, Xiaobo An, and Joseph Dunlop. *Demographic Effects on Estimates of Automatic Face Recognition Performance*. n.d.

Pound, Roscoe. "The Limits of Effective Legal Action." *The International Journal of Ethics* 27, no. 2 (1917): 150–67. https://doi.org/10.1086/intejethi.27.2.2377288.

Raghavan, Manish. *The Societal Impacts of Algorithmic Decision-Making*. 1st ed. ACM, 2023. https://doi.org/10.1145/3603195.

Raghavan, Manish. *The Societal Impacts of Algorithmic Decision-Making*. 1st ed. ACM, 2023. https://doi.org/10.1145/3603195.

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying down Harmonised Rules on Artificial Intelligence and Amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No

168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA Relevance) (2024). http://data.europa.eu/eli/reg/2024/1689/oj.

Schwarting, Rena, and Lena Ulbricht. "Why Organization Matters in 'Algorithmic Discrimination.'" *KZfSS Kölner Zeitschrift Für Soziologie Und Sozialpsychologie* 74, no. S1 (2022): 307–30. https://doi.org/10.1007/s11577-022-00838-3.

Selbst, Andrew D., Danah Boyd, Sorelle A. Friedler, Suresh Venkatasubramanian, and Janet Vertesi. "Fairness and Abstraction in Sociotechnical Systems." *Proceedings of the Conference on Fairness, Accountability, and Transparency* (New York, NY, USA), FAT* '19, January 29, 2019, 59–68. https://doi.org/10.1145/3287560.3287598.

Wang, Hao. "Transparency as Manipulation? Uncovering the Disciplinary Power of Algorithmic Transparency." *Philosophy & Technology* 35, no. 3 (2022): 69. https://doi.org/10.1007/s13347-022-00564-w.

Yeung, Karen. "Algorithmic Regulation: A Critical Interrogation." *Regulation & Governance* 12, no. 4 (2018): 505–23. https://doi.org/10.1111/rego.12158.