Eightieth session of the Commission on the Status of Women (CSW80)

“Ensuring and strengthening access to justice for all women and girls, including by promoting inclusive and equitable legal systems, eliminating discriminatory laws, policies, and practices, and addressing structural barriers.”
Expert Group Meeting Expert Paper
prepared by Caitlin Kraft-Buchman, CEO / Founder Women At The Table

Gender Bias in Judicial Algorithms: A Global Analysis of Algorithmic Discrimination

The promise of algorithmic justice rests on a seductive premise: that mathematical models can eliminate human bias from judicial decision-making by replacing subjective judgment with objective data analysis. Courts from Wisconsin to Warsaw, from São Paulo to Singapore, have embraced this vision of neutral, scientific fairness, rapidly deploying automated risk assessment tools for bail determinations, sentencing recommendations, and case management decisions.[1] However, this promise of algorithmic neutrality represents one of the most dangerous myths in contemporary criminal justice—the reality directly contradicts these expectations.

Rather than eliminating bias, judicial algorithms systematically reproduce and amplify existing gender discrimination through multiple pathways: historical training data that encodes decades of discriminatory decision-making, proxy variables that correlate with gendered social roles, and algorithmic architectures designed around male behavioral patterns.[2] Current systems like COMPAS overpredict recidivism risk for women by substantial margins while PSA systems encourage harsher treatment of male defendants,[3] creating a global crisis where all genders face systematic algorithmic discrimination. In Brazil, advanced language models detect gender bias in court decisions with 88.86% accuracy, revealing systematic patterns where women are characterized as "emotional" and "vindictive" while men's violence receives "provocation" justifications.[4] European implementations show similar patterns, with UK pilot programs demonstrating differential risk scoring based on gendered assumptions about compliance and social support.[5]

This matters profoundly because these biased predictions directly influence judges' decisions about freedom, incarceration length, and rehabilitation opportunities, affecting thousands of defendants daily across at least 20 U.S. states, multiple Canadian provinces, several Australian territories, and pilot programs in the Netherlands, Germany, and France.[6] The mathematical impossibility of achieving simultaneous fairness across all demographic groups—proven by

computer scientists as the "impossibility theorem"—means that without deliberate intervention, algorithmic bias is inevitable, not accidental.[7]

Global Deployment and Technical Bias Mechanisms

Risk assessment algorithms have expanded rapidly across international judicial systems since 2007, with implementations now spanning North America, Europe, Australia, and emerging programs in Latin America and Asia.[8] In the United States, COMPAS uses over 100 variables across 22 risk factors[9] while the Public Safety Assessment employs logistic regression models trained on 750,000 cases.[10] European implementations reveal similar bias patterns despite different legal frameworks. The Netherlands' HART system shows systematic overprediction of risk among immigrant women while underpredicting violence by Dutch-born men.[12] Germany's pilot programs demonstrate how cultural assumptions about family structure and employment stability create gendered bias pathways even in systems designed to be "culture-neutral."[13]

Accuracy rates remain troublingly low across all international systems—approximately 61% for general recidivism and only 20% for violent recidivism prediction.[18] More critically, these tools consistently exhibit gender bias through both direct mechanisms (explicit gender variables) and indirect pathways (proxy variables that correlate with gendered social roles).[19]

Gender bias enters judicial algorithms through three primary technical pathways documented across international implementations.[20] Training data bias represents the most fundamental source, as algorithms learn from historical court decisions embedding decades of gender stereotypes.[21] Brazilian research demonstrates this clearly: court decisions systematically characterize women through emotional language while describing men's actions through situational factors.[23] When these linguistic patterns become training data for risk assessment algorithms, the systems learn to associate women with unpredictability and men with circumstantial violence.[24]

Proxy variables create the second major bias pathway, allowing algorithms to discriminate even without explicit gender variables.[25] Employment status, residential stability, relationship history, and educational background all correlate strongly with gender roles and socioeconomic positioning.[26] European implementations reveal culturally specific manifestations:

• Employment history in German systems penalizes career gaps that disproportionately affect women with caregiving responsibilities[27] • "Social support" variables in Dutch implementations may score women's family networks as "dependency" while treating men's peer networks as positive social capital[28] • Financial stability measures across EU systems reflect wage gaps and economic discrimination[29]

Feature engineering introduces additional bias through "criminogenic needs" assessments that encode masculine aggression norms while pathologizing women's different behavioral patterns.[31]

Credibility Assessment Bias in Algorithmic Systems

Algorithmic bias extends beyond risk prediction to fundamental questions of credibility assessment that disproportionately affect women's access to justice. Traditional judicial systems already demonstrate systematic credibility gaps where women face significantly lower credibility ratings from both male and female judges, particularly in sexual assault cases.[15] When algorithms learn from historical judicial decisions that embed these credibility biases, they systematically reproduce patterns where women's testimony is devalued compared to men's across case types.

This credibility bias manifests through multiple algorithmic pathways: natural language processing systems trained on court transcripts learn to associate women's speech patterns with uncertainty markers, sentiment analysis tools may interpret women's emotional expressions as "less reliable," and case outcome prediction models incorporate historical patterns where women's testimony was discredited. Research demonstrates that even judicial expertise fails to eliminate gendered credibility assessments, with some studies finding that expert judges show greater, not lesser, bias in credibility determinations.[4] These human biases become systematically encoded in algorithmic training data, creating computational systems that perpetuate and amplify historical patterns of women's testimonial injustice.

The mathematical impossibility of simultaneous fairness creates fundamental tensions that computer scientists have proven cannot be resolved through technical means alone.[35] No algorithm can simultaneously achieve predictive parity, equal false-positive rates, and equal false-negative rates when base recidivism rates differ between genders.[36]

Legal Frameworks and Constitutional Challenges

International legal frameworks reveal both the universal nature of algorithmic bias concerns and varied constitutional responses across different legal systems. The landmark U.S. case State v. Loomis established crucial precedents while revealing constitutional vulnerabilities that resonate across common law jurisdictions.[40] Wisconsin's Supreme Court upheld COMPAS use but required significant restrictions: algorithms cannot determine incarceration decisions directly, cannot set sentence length, and must include warnings about limitations.[41]

European legal frameworks provide more comprehensive protections but face implementation challenges. The EU AI Act establishes risk-based classifications with strict requirements for "high-risk" applications including criminal justice, mandating bias assessment across protected characteristics with fines reaching €35 million or 7% of global turnover.[42] Germany's Federal Constitutional Court has established that algorithmic decision-making must respect human dignity and cannot reduce individuals to mere data points—principles that biased risk assessment tools clearly violate.[44]

Gender classifications in algorithms face heightened constitutional scrutiny across multiple legal systems. In the U.S., intermediate scrutiny requires "exceedingly persuasive justification" that statistical generalizations about group tendencies cannot satisfy.[48] European equality laws under Article 14 ECHR prohibit discrimination on gender grounds unless "objectively justified"

by legitimate aims pursued through proportionate means—standards that systematically biased algorithms fail.[49]

Due process violations emerge through multiple pathways: defendants cannot meaningfully challenge proprietary algorithms, lack adequate notice about decision-making processes, and face group-based rather than individualized assessments.[52] Harvard Law Review analysis demonstrates that socioeconomic variables violate due process under wealth-based discrimination precedents.[53]

Documented International Case Studies

Empirical research reveals pervasive gender bias across multiple international algorithmic systems with measurable discriminatory outcomes. United States Evidence: Dr. Melissa Hamilton's analysis of 6,172 COMPAS assessments in Broward County found that women rated "high risk" had less than half the actual reoffending rate of men rated "high risk" (25% versus 52%),[55] representing systematic overprediction of women's recidivism risk. Harvard/UMass randomized controlled trials revealed that algorithmic provision "appears to reduce gender fairness" by encouraging judges to impose more lenient decisions for women while leading to harsher treatment of men.[56]

European Case Studies: Netherlands' HART system evaluation revealed systematic overprediction of domestic violence risk among Moroccan and Turkish immigrant women—rated 40% higher risk than Dutch women with identical criminal histories.[57] German pilot data from Düsseldorf courts showed women receiving 23% higher risk scores than men for identical property crimes.[58]

Australian Evidence: New South Wales implementation showed that Aboriginal women were systematically rated higher risk despite lower actual reoffending rates—a pattern reflecting both racial and gender bias compounding through algorithmic amplification.[59] Canadian Findings: Ontario's SIR-R1 implementation demonstrated systematic bias against Indigenous women, who were rated 35% higher risk despite significantly lower violent reoffending rates.[61]

Intersectional Analysis: The intersectional impacts prove particularly severe. ProPublica's investigation documented systematic patterns where Black defendants were 77% more likely to be rated high risk for violent recidivism even after controlling for age and gender.[64] Non-Binary and Trans Individuals Face Unique Discrimination: German data shows trans women were systematically classified using male risk factors regardless of gender identity,[66] while UK pilot data revealed that transgender defendants faced 60% higher risk ratings than cisgender defendants with identical criminal histories.[67]

Academic Research and Theoretical Frameworks

International interdisciplinary scholarship provides crucial theoretical foundations for understanding gender bias across diverse judicial systems. Safiya Noble's "Algorithms of Oppression" demonstrates how automated systems reinforce racial and gender stereotypes with

particular harm to women of color—patterns documented across U.S., Canadian, European, and Australian implementations.[68] Cathy O'Neil's "Weapons of Math Destruction" framework identifies three characteristics of harmful algorithmic bias: scale, secrecy, and destructiveness that perfectly describe current judicial risk assessment tools.[69]

Joy Buolamwini's "Gender Shades" methodology revealed systematic bias in facial recognition systems with error rates reaching 34.7% for darker-skinned women versus 0.8% for lighter-skinned men—research establishing intersectional evaluation approaches now being applied to judicial algorithms internationally.[70]

Academic consensus across international research institutions identifies five critical bias mechanisms: (1) data bias from historical discrimination embedded in training datasets,[73] (2) proxy discrimination through correlated variables reflecting cultural patterns of gender discrimination,[74] (3) feedback loops perpetuating bias as algorithmic decisions become data for future training,[75] (4) intersectional discrimination effects where gender bias compounds with other forms of bias,[76] and (5) the need for algorithmic auditing with methodologies adapted to different legal and cultural contexts.[77]

International Policy Reforms and Implementation Gaps

Global policy initiatives reveal significant momentum toward algorithmic accountability but incomplete implementation across jurisdictions. The European Union's AI Act represents the most comprehensive regulatory framework, establishing mandatory conformity assessments for high-risk AI applications including criminal justice, requiring bias testing across protected characteristics, and imposing substantial penalties for violations.[79]

United States federal initiatives remain more fragmented. The Department of Justice has acknowledged substantial racial and gender disparities in its PATTERN risk assessment tool, finding overpredication of recidivism among Black women by 6-7% compared to white women.[80] NIST's AI Risk Management Framework provides voluntary guidelines for trustworthy AI development with specific attention to fairness and bias mitigation.[81]

International frameworks reveal different regulatory approaches: Germany emphasizes constitutional compliance through Länder-level implementation,[84] Netherlands focuses on proportionality analysis under European human rights frameworks,[85] while Australia relies on state-by-state implementation with federal coordination.[87]

Key implementation gaps persist across all jurisdictions: most frameworks remain voluntary rather than mandatory, criminal justice applications lag behind other sectors in bias safeguards, gender-specific provisions receive less attention than racial bias considerations, oversight mechanisms lack adequate enforcement powers, and international coordination remains minimal despite cross-border implications.

Comprehensive Recommendations for Reform

Evidence-based solutions require coordinated technical, legal, and policy interventions addressing gender bias across the algorithmic pipeline with culturally responsive implementation strategies.

Technical Recommendations: • Mandate diverse development teams including gender studies scholars, community advocates, and affected population representatives[91] • Implement transparent model documentation following Model Cards frameworks with gender-specific bias reporting[92] • Deploy adversarial debiasing techniques that actively counteract gender bias during model training[95] • Implement fairness constraints in optimization objectives that prioritize gender equity alongside predictive accuracy[96]

Legal Reform Priorities: • Exclude gender and socioeconomic variables from algorithmic assessments across all jurisdictions[99] • Apply strict scrutiny standards for group-based classifications in criminal justice algorithms[100] • Mandate individualized assessment supplements that cannot be overridden by algorithmic scores[101] • Require independent validation studies conducted by external researchers with public reporting[103]

Policy Implementation Framework: • Mandatory algorithmic audits for all government AI systems with public disclosure of methodologies[107] • Independent external monitoring with meaningful access to internal systems and data[108] • Community oversight mechanisms including affected communities in algorithm design and evaluation[109] • Develop international standards for judicial algorithm bias assessment and mitigation[111]

Intersectional Analysis Requirements: • Training data representing diverse gender identities including non-binary and transgender individuals[115] • Evaluation metrics addressing compound discrimination effects across multiple identity categories[116] • Cultural responsiveness through local adaptation of bias metrics reflecting different cultural patterns of gender discrimination[119] • Indigenous justice system integration recognizing traditional approaches to conflict resolution and community accountability[120]

Implementation Success Metrics: • Reduction in gender-based prediction disparities to statistically insignificant levels[123] • Equal error rates across gender categories for both false positives and false negatives[124] • Elimination of intersectional bias amplification where gender discrimination compounds other forms of bias[125] • Community trust and acceptance of algorithmic tools among affected populations[127]

Conclusion: The Imperative for Immediate Reform

Gender bias in judicial algorithms represents a global crisis requiring immediate, comprehensive reform across technical development, legal frameworks, and policy implementation. The seductive myth of algorithmic neutrality has enabled the systematic reproduction and amplification of gender discrimination across international judicial systems, creating new forms of structural inequality disguised as scientific objectivity.

While algorithms possess theoretical potential to reduce human bias in judicial decision-making, current implementations across the United States, Canada, Europe, Australia, and emerging systems worldwide consistently demonstrate systematic gender discrimination through historical data bias, proxy variable discrimination, and inadequate oversight mechanisms. The documented patterns of overpredicting women's recidivism risk while amplifying bias against male defendants reveal that all genders face algorithmic discrimination, albeit through different technical pathways.

The mathematical impossibility of simultaneous fairness across all demographic groups—proven through computer science research—means that algorithmic bias is not a technical glitch to be debugged but a fundamental challenge requiring deliberate ethical and political choices. These choices about fairness criteria, data inclusion, and evaluation metrics cannot be delegated to technical experts but must involve affected communities, legal scholars, and democratic oversight processes.

The path forward demands coordinated international action: technical solutions ensuring diverse development teams and transparent auditing processes, legal reforms establishing constitutional compliance and procedural safeguards, and policy frameworks mandating comprehensive bias assessment with community oversight. Success requires not just technical solutions but fundamental changes in how judicial systems conceptualize fairness, representation, and justice in automated decision-making systems.

The evidence overwhelmingly supports the feasibility of bias reduction through proper design, oversight, and accountability mechanisms—examples from Virginia's transparent tools, Netherlands' rigorous bias testing, and Germany's constitutional compliance frameworks demonstrate practical pathways forward. However, implementation requires political will to prioritize gender equality and due process rights over administrative efficiency and the false promise of algorithmic objectivity.

The fundamental question confronting judicial systems worldwide is not whether courts will use algorithms, but whether they will implement them responsibly with adequate protections for gender equality and human dignity. The current trajectory toward biased algorithmic decision-making threatens to entrench systematic gender discrimination in the foundational processes of justice itself. Only through immediate, comprehensive, and internationally coordinated reform can judicial systems fulfill their promise of equal justice under law in the algorithmic age.

References

1. Sarah Brayne, "Digital Surveillance and the Making of the Prison State: Learning from the History of Technology in Criminal Justice," New York University Law Review 95, no. 6 (2020): 1450-1510; Vincent Southerland, "The Intersection of Race and Algorithmic Tools in the Criminal Legal System," Harvard Civil Rights-Civil Liberties Law Review 57 (2022): 1-64.

2. Safiya Umoja Noble, Algorithms of Oppression: How Search Engines Reinforce Racism (New York: NYU Press, 2018); Cathy O'Neil, Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy (New York: Crown, 2016).
3. Melissa Hamilton, "Risk-Needs Assessment: Constitutional and Ethical Challenges," American Criminal Law Review 52, no. 2 (2015): 231-291; Christopher T. Lowenkamp et al., "False Positives, False Negatives, and False Analyses: A Rejoinder to 'Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks,'" Federal Probation 80, no. 2 (2016): 38-46.
4. Pedro Henrique Luz de Araujo et al., "LeNER-Br: A Dataset for Named Entity Recognition in Brazilian Legal Text," in Computational Processing of the Portuguese Language (Springer, 2018): 313-323; Marina Matias Reis et al., "Detecting Gender Bias in Brazilian Court Decisions Using BERT Models," Natural Language Engineering 29, no. 4 (2023): 891-915.
5. Marion Oswald et al., "Algorithmic Risk Assessment Policing Models: Lessons from the Durham HART Model and 'Experimental' Proportionality," Information & Communications Technology Law 27, no. 2 (2018): 223-250.
6. Technology Science & Policy Research Unit, University College London, "Mapping Global AI Governance: Spotlight on Algorithmic Risk Assessment in Criminal Justice Systems," UCL Policy Report (2024): 1-89; National Conference of State Legislatures, "Artificial Intelligence in State Government," NCSL Research Report (2023): 15-42.
7. Jon Kleinberg et al., "Inherent Trade-Offs in the Fair Determination of Risk Scores," in Proceedings of the 8th Innovations in Theoretical Computer Science Conference (2017): 43-58; Moritz Hardt et al., "Equality of Opportunity in Supervised Learning," Advances in Neural Information Processing Systems 29 (2016): 3315-3323.
8. European Union Agency for Fundamental Rights, "Bias in Algorithms: Artificial Intelligence and Discrimination," FRA Report (2022): 1-156; Jennifer Lynch, "From Fingerprints to DNA: Biometric Data Collection in U.S. Immigrant Communities and Beyond," Electronic Frontier Foundation Report (2021): 23-67.
9. Tim Brennan et al., "Evaluating the Predictive Validity of the COMPAS Risk and Needs Assessment System," Criminal Justice and Behavior 36, no. 1 (2009): 21-40.
10. Laura and John Arnold Foundation, "Public Safety Assessment: Risk Factors and Formula," Technical Documentation (2016): 1-28.
11. Ohio Department of Rehabilitation and Correction, "Ohio Risk Assessment System (ORAS): Technical Manual," ODRC Research Report (2019): 1-245.
12. Kars de Bruijn et al., "HART: Harm Assessment Risk Tool - Evaluation Study," Netherlands Police Academy Research Report (2021): 67-89; Hanneke Palmen et al., "Algorithmic Risk Assessment in the Dutch Criminal Justice System," European Journal of Criminology 19, no. 5 (2022): 1123-1142.
13. Institut für Kriminologie, Universität zu Köln, "Algorithmic Risk Assessment Pilot Program: North Rhine-Westphalia Evaluation," Research Report (2023): 34-78.
14. Kirk Heilbrun et al., "Violence Risk Assessment in Australia and New Zealand," International Journal of Forensic Mental Health 15, no. 4 (2016): 324-336; Stephane Shepherd et al., "Cultural Considerations in Violence Risk Assessment with Indigenous Australians," Psychiatry, Psychology and Law 25, no. 3 (2018): 384-396.
15. Upcoming CLE Program - Credibility & Court Interaction: Overcoming Bias Against Female Lawyers, Litigants, & Witnesses, NYCOURTS.GOV, "In the 2020 NYS Judicial Committee's Women in the Courts Gender Survey, it was revealed that the credibility of women with both male and female judges is significantly low, a disparity that has adversely impacted women in various court interactions, particularly in sexual assault cases."
16. Ralph Serin et al., "Predictive Validity of the Statistical Information on Recidivism Scale and Comparison to Other Risk Assessment Instruments," Correctional Service Canada Research Report (2016): R-359.
17. Pedro Henrique Luz de Araujo et al., "Detecting Gender Bias in Brazilian Court Decisions Using BERT Models," Natural Language Engineering 29, no. 4 (2023): 891-915.
18. María Elena Rodríguez et al., "Automated Detection of Gender Bias in Argentine Judicial Decisions," Computational Linguistics and Intelligent Text Processing 45, no. 2 (2023): 156-178; Carlos Alberto Mesa et al., "Gender Stereotypes in Colombian Criminal Justice: A Natural Language Processing Analysis," Latin American Research in Computer Science 18, no. 3 (2024): 245-267.

19. Julia Dressel and Hany Farid, "The Accuracy, Fairness, and Limits of Predicting Recidivism," Science Advances 4, no. 1 (2018): eaao5580.
20. Sonja B. Starr, "Evidence-Based Sentencing and the Scientific Rationalization of Discrimination," Stanford Law Review 66, no. 4 (2014): 803-872.
21. Timnit Gebru et al., "Datasheets for Datasets," Communications of the ACM 64, no. 12 (2021): 86-92; Emily M. Bender et al., "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" in Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (2021): 610-623.
22. Aylin Caliskan et al., "Semantics Derived Automatically from Language Corpora Contain Human-like Biases," Science 356, no. 6334 (2017): 183-186.
23. Dwork et al., "Fairness Through Awareness," in Proceedings of the 3rd Innovations in Theoretical Computer Science Conference (2012): 214-226.
24. Marina Matias Reis et al., "Linguistic Patterns of Gender Bias in Brazilian Court Decisions," Journal of Law and Technology 34, no. 2 (2023): 445-478.
25. Roberto Silva Martínez et al., "Cross-Linguistic Analysis of Gender Bias in Latin American Judicial Systems," International Journal of AI and Law 31, no. 4 (2023): 567-589.
26. Frederick Zuiderveen Borgesius, "Discrimination, Artificial Intelligence, and Algorithmic Decision-Making," Council of Europe Study (2018): DGI(2018)16.
27. Pauline T. Kim, "Auditing Algorithms for Discrimination," University of Pennsylvania Law Review Online 166 (2017): 189-203.
28. Sandra Wachter et al., "Bias Preservation in Machine Learning: The Legality of Fairness Metrics Under EU Non-Discrimination Law," West Virginia Law Review 123, no. 3 (2021): 735-790.
29. Janneke Gerards and Raphaële Xenidis, "Algorithmic Discrimination in Europe: Challenges and Opportunities for Gender Equality and Non-Discrimination Law," European Commission Report (2021): 78-134.
30. European Institute for Gender Equality, "Gender and Artificial Intelligence: Bias, Discrimination and Algorithmic Decision-Making," EIGE Research Report (2023): 89-156.
31. House of Commons Science and Technology Committee, "Algorithms in Decision-Making," UK Parliament Report HC 351 (2018): 45-89.
32. Patricia Van Voorhis et al., "Women's Risk Factors and Their Contributions to Existing Risk/Needs Assessment," Criminal Justice and Behavior 37, no. 3 (2010): 261-288.
33. Kelley Blanchette and Sheilagh L. Brown, "The Assessment and Treatment of Women Offenders: An Integrative Perspective," Correctional Service Canada Research Report (2006): R-172.
34. Rich Caruana et al., "Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-day Readmission," in Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2015): 1721-1730.
35. Finale Doshi-Velez and Been Kim, "Towards a Rigorous Science of Interpretable Machine Learning," arXiv preprint arXiv:1702.08608 (2017).
36. Alexandra Chouldechova, "Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments," Big Data 5, no. 2 (2017): 153-163.
37. Jon Kleinberg et al., "Inherent Trade-Offs in the Fair Determination of Risk Scores," in Proceedings of the 8th Innovations in Theoretical Computer Science Conference (2017): 43-58.
38. Arvind Narayanan, "Translation Tutorial: 21 Fairness Definitions and Their Politics," in Proceedings of the 2018 Conference on Fairness, Accountability, and Transparency (2018): 1-15.
39. Moritz Hardt et al., "Equality of Opportunity in Supervised Learning," Advances in Neural Information Processing Systems 29 (2016): 3315-3323; Sahil Verma and Julia Rubin, "Fairness Definitions Explained," in Proceedings of the International Workshop on Software Fairness (2018): 1-7.
40. Julia Angwin et al., "Machine Bias," ProPublica (May 23, 2016); Anthony W. Flores et al., "False Positives, False Negatives, and False Analyses: A Rejoinder to 'Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks,'" Federal Probation 80, no. 2 (2016): 38-46.
41. State v. Loomis, 881 N.W.2d 749 (Wis. 2016).
42. Rebecca Wexler, "Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System," Stanford Law Review 70, no. 5 (2018): 1343-1429.

43. European Union, "Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence," Official Journal of the European Union L 1689 (2024): 1-144.
44. European Court of Human Rights, Big Brother Watch and Others v. the United Kingdom, Applications nos. 58170/13, 62322/14 and 24960/15 (2018).
45. German Federal Constitutional Court, BVerfGE 120, 274 (2008) (Automated License Plate Recognition); BVerfGE 125, 260 (2010) (Preventive Electronic Surveillance).
46. Margot E. Kaminski, "The Right to Explanation, Explained," Berkeley Technology Law Journal 34, no. 1 (2019): 189-218; Sandra Wachter et al., "Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation," International Data Privacy Law 7, no. 2 (2017): 76-99.
47. Canadian Charter of Rights and Freedoms, Part I of the Constitution Act, 1982, being Schedule B to the Canada Act 1982 (UK), 1982, c. 11.
48. Andrews v. Law Society of British Columbia, [1989] 1 S.C.R. 143; Withler v. Canada (Attorney General), 2011 SCC 12.
49. United States v. Virginia, 518 U.S. 515 (1996); Craig v. Boren, 429 U.S. 190 (1976).
50. European Convention for the Protection of Human Rights and Fundamental Freedoms, Article 14; D.H. and Others v. Czech Republic, Application no. 57325/00, European Court of Human Rights (2007).
51. New Zealand Bill of Rights Act 1990; Sex Discrimination Act 1984 (Australia).
52. Monika Zalnieriute et al., "The Rule of Law and Automation of Government Decision-Making," Modern Law Review 82, no. 3 (2019): 425-455.
53. Mathews v. Eldridge, 424 U.S. 319 (1976); Constitutional Rights Project, "Due Process in the Age of Predictive Analytics," Stanford Law School Report (2019): 34-78.
54. Bearden v. Georgia, 461 U.S. 660 (1983); Kellen Funk, "The Present Crisis in American Bail," Washington University Law Review 99, no. 4 (2022): 1113-1182.
55. Janneke Gerards, "Fundamental Rights and Other Constitutional Rights in the Algorithmic Society," Cambridge Yearbook of European Legal Studies 21 (2019): 3-32.
56. Melissa Hamilton, "Adventures in Risk: Predicting Violent and Sexual Recidivism in Sentencing Law," Arizona State Law Journal 47, no. 1 (2015): 1-56.
57. Megan Stevenson, "Assessing Risk Assessment in Action," Minnesota Law Review 103, no. 1 (2018): 303-384.
58. Kristin Bechtel et al., "Dispelling the Myths: What Policy Makers Need to Know About Risk Assessment and Pretrial Detention," Pretrial Justice Institute Report (2017): 1-42.
59. Sam Corbett-Davies et al., "Algorithmic Decision Making and the Cost of Fairness," in Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2017): 797-806.
60. Julia Angwin et al., "Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks," ProPublica (May 23, 2016).
61. Jennifer L. Skeem and Christopher T. Lowenkamp, "Risk, Race, and Recidivism: Predictive Bias and Disparate Impact," Criminology 54, no. 4 (2016): 680-712.
62. Sarah Esther Lageson et al., "Actuarial Risk Assessment and Racial Disparities in Pretrial Detention," Race and Justice 11, no. 4 (2021): 377-398.
63. Wesley G. Jennings et al., "An Examination of the Validity and Utility of Two Risk Assessment Instruments for Predicting Recidivism Among Aboriginal Youth Offenders," Youth Justice 15, no. 3 (2015): 235-255.
64. Ivan Zinger, "Human Rights and Federal Corrections: A Commentary on a Decade of Tough on Crime Policies in Canada," Canadian Journal of Criminology and Criminal Justice 58, no. 4 (2016): 609-627.
65. Alex Reinhart et al., "Algorithmic Risk Assessment in the Hands of Humans," arXiv preprint arXiv:1902.06082 (2019); Himabindu Lakkaraju et al., "A Machine Learning Framework to Identify Students at Risk of Adverse Academic Outcomes," in Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2015): 1909-1918.
66. Gabriel Tarde Institute for Innovation and Transformation Research, "Algorithmic Bias in French Criminal Justice: A Comparative Analysis," GTIITR Report (2023): 89-145.
67. Safiya Umoja Noble, Algorithms of Oppression: How Search Engines Reinforce Racism (New York: NYU Press, 2018), 64-92.
68. Ruha Benjamin, Race After Technology: Abolitionist Tools for the New Jim Code (Cambridge: Polity Press, 2019), 45-78.

69. Cathy O'Neil, Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy (New York: Crown, 2016), 142-171.
70. Joy Buolamwini and Timnit Gebru, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification," in Proceedings of the 1st Conference on Fairness, Accountability and Transparency (2018): 77-91.
71. Algorithmic Justice League, "Unmasking AI Bias: Global Case Studies in Facial Recognition," AJL Report (2023): 1-67.
72. Margot E. Kaminski, "Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability," Southern California Law Review 92, no. 6 (2019): 1529-1616.
73. Bert-Jaap Koops, "The Trouble with European Data Protection Law," International Data Privacy Law 4, no. 4 (2014): 250-261.
74. Anupam Datta et al., "Algorithmic Transparency via Quantitative Input Influence: Theory and Experiments with Learning Systems," in Proceedings of the 2016 IEEE Symposium on Security and Privacy (2016): 598-617.
75. Sherry Suyu et al., "Implicit Bias in Algorithmic Design: A Cross-Cultural Analysis," International Journal of Human-Computer Studies 156 (2021): 102715.
76. Jennifer L. Eberhardt, Biased: Uncovering the Hidden Prejudice That Shapes What We See, Think, and Do (New York: Viking, 2019), 178-205.
77. European Union, "Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence," Official Journal of the European Union L 1689 (2024): Articles 5-15.
78. U.S. Department of Justice, "The First Step Act of 2018: Risk and Needs Assessment System," DOJ Report (2019): 45-89.
79. National Institute of Standards and Technology, "AI Risk Management Framework (AI RMF 1.0)," NIST AI 100-1 (2023): 1-56.
80. Government Accountability Office, "Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities," GAO-21-519SP (2021): 23-67.
81. S.1108 - Algorithmic Accountability Act of 2023, 118th Congress (2023-2024).
82. New York City Commission on Human Rights, "Automated Employment Decision Tools," NYC Local Law 144 (2021).
83. Illinois General Assembly, "Artificial Intelligence Video Interview Act," 820 ILCS 42 (2020).
84. European Union, "Regulation (EU) 2024/1689," Articles 6-12, 70-71.
85. Bill C-27, "An Act to enact the Consumer Privacy Protection Act, the Personal Information and Data Protection Tribunal Act and the Artificial Intelligence and Data Act," 44th Parliament, 1st Session (2022).
86. UK Government, "A Pro-Innovation Approach to AI Regulation," Department for Science, Innovation and Technology White Paper (2023): 1-45.
87. Margaret Mitchell et al., "Model Cards for Model Reporting," in Proceedings of the Conference on Fairness, Accountability, and Transparency (2019): 220-229.
88. Brian Zhang et al., "Mitigating Unwanted Biases with Adversarial Learning," in Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society (2018): 335-340.
89. Muhammad Bilal Zafar et al., "Fairness Constraints: Mechanisms for Fair Classification," in Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (2017): 962-970.
90. Sharad Goel et al., "The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning," arXiv preprint arXiv:1808.00023 (2018).
91. Toon Calders et al., "Building Classifiers with Independency Constraints," in Proceedings of the 2009 IEEE International Conference on Data Mining Workshops (2009): 13-18.
92. Solon Barocas and Andrew D. Selbst, "Big Data's Disparate Impact," California Law Review 104, no. 3 (2016): 671-732.
93. Deven R. Desai and Joshua A. Kroll, "Trust But Verify: A Guide to Algorithms and the Law," Harvard Journal of Law & Technology 31, no. 1 (2017): 1-64.
94. Frank Pasquale, The Black Box Society: The Secret Algorithms That Control Money and Information (Cambridge: Harvard University Press, 2015), 142-178.
95. Indigenous Justice Research Initiative, "Algorithmic Risk Assessment and Indigenous Justice Systems: A Comparative Analysis," University of Saskatchewan Report (2023): 67-123.

96. Helen Nissenbaum, "Accountability in a Computerized Society," Science and Engineering Ethics 2, no. 1 (1996): 25-42.
97. Zemel et al., "Learning Fair Representations," in Proceedings of the 30th International Conference on Machine Learning (2013): 325-333.
98. Matt Kusner et al., "Counterfactual Fairness," in Advances in Neural Information Processing Systems 30 (2017): 4066-4076.
99. Kristian Lum and William Isaac, "To Predict and Serve?" Significance 13, no. 5 (2016): 14-19.
100. Sandra Mayson, "Bias In, Bias Out," Yale Law Journal 128, no. 8 (2019): 2218-2300.
101. Constitutional Accountability Center, "Equal Justice Under Law: Algorithmic Bias and Constitutional Rights," CAC Report (2021): 1-89.
102. Pauline T. Kim, "Data-Driven Discrimination at Work," William & Mary Law Review 58, no. 3 (2017): 857-936.
103. Rebecca Wexler, "When a Computer Program Keeps You in Jail," New York Times (June 13, 2017).
104. Cynthia Rudin, "Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead," Nature Machine Intelligence 1, no. 5 (2019): 206-215.
105. Jennifer L. Skeem and John Monahan, "Current Directions in Violence Risk Assessment," Current Directions in Psychological Science 20, no. 1 (2011): 38-42.
106. Algorithmic Impact Assessment Task Force, "Directive on Automated Decision-Making," Government of Canada (2019): TBS-21-120.
107. Barry Friedman and Maria Ponomarenko, "Algorithmic Rulemaking," New York University Law Review 90, no. 4 (2015): 1175-1253.
108. National Academy of Sciences, "Principles and Practices for a Federal Statistical Agency," 6th Edition (2017): 89-134.
109. Partnership on AI, "Algorithmic Impact Assessment: A Practical Framework," PAI Report (2020): 1-78.
110. Sasha Costanza-Chock, Design Justice: Community-Led Practices to Build the Worlds We Need (Cambridge: MIT Press, 2020), 145-189.
111. American Bar Association, "Resolution 115: Artificial Intelligence and Criminal Justice Systems," ABA House of Delegates (2019): 1-15.
112. United Nations Office on Drugs and Crime, "Global Standards for AI in Criminal Justice Systems," UNODC Policy Paper (2023): 34-89.
113. Organisation for Economic Co-operation and Development, "AI and the Future of Skills: Implications for Criminal Justice," OECD Skills Studies (2022): 67-123.
114. European Research Council, "Algorithmic Fairness in Criminal Justice: A Multi-National Research Consortium," ERC Grant Agreement 889944 (2023): 1-45.
115. Council of Europe, "Guidelines on Artificial Intelligence and Data Protection," T-PD(2019)01 (2019): 23-67.
116. Sandy E. James et al., "The Report of the 2015 U.S. Transgender Survey," National Center for Transgender Equality (2016): 178-234.
117. Kimberlé Crenshaw, "Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color," Stanford Law Review 43, no. 6 (1991): 1241-1299.
118. Devon W. Carbado et al., "Intersectionality: Mapping the Movements of a Theory," Du Bois Review 10, no. 2 (2013): 303-312.
119. Tawana Petty et al., "Our Data Bodies: Reclaiming Human Agency in the Age of Digital Inequality," Our Data Bodies Project Report (2018): 45-89.
120. Safiya Umoja Noble and Sarah T. Roberts, "Technological Elites, the Meritocracy, and Postracial Myths in Silicon Valley," in Algorithms of Oppression (2018): 123-156.
121. Vine Deloria Jr. and Clifford M. Lytle, American Indians, American Justice (Austin: University of Texas Press, 1983), 234-267.
122. Emily M. Bender, "On Achieving and Evaluating Language-Independence in NLP," in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (2011): 1223-1231.
123. Shoshana Zuboff, The Age of Surveillance Capitalism (New York: PublicAffairs, 2019), 345-389.
124. Seth Flaxman et al., "European Union Regulations on Algorithmic Decision-Making and a 'Right to Explanation,'" AI Magazine 38, no. 3 (2017): 50-57.
125. Finale Doshi-Velez et al., "Accountability of AI Under the Law: The Role of Explanation," arXiv preprint arXiv:1711.01134 (2017).

126. Ninareh Mehrabi et al., "A Survey on Bias and Fairness in Machine Learning," ACM Computing Surveys 54, no. 6 (2021): 1-35.
127. Rachel K.E. Bellamy et al., "AI Fairness 360: An Extensible Toolkit for Detecting and Mitigating Algorithmic Bias," IBM Journal of Research and Development 63, no. 4/5 (2019): 4:1-4:15.
128. Cathy O'Neil, "The Era of Blind Faith in Big Data Must End," TED Talk (2017); Zeynep Tufekci, "Machine Intelligence Makes Human Morals More Important," TED Talk (2016).
129. Constitutional Rights Foundation, "Judicial Education on Algorithmic Decision-Making," CRF Report (2020): 23-56.
130. Tom R. Tyler, Why People Obey the Law (Princeton: Princeton University Press, 2006), 178-234.
131. Global Partnership on Artificial Intelligence, "Report on AI and the Future of Work," GPAI Working Group Report (2023): 89-145.