ISSN(Print) : 3105-7357
ISSN (Online) : 3105-7365

RESEARCH ARTICLE

Social Thought and Policy Review

Volume: 04 Issue: 01(2026)

OPEN ACCESS

Predictive Policing, Bias, and Community Legitimacy: A Social Justice Evaluation Framework

[1] **Ayesha Mahmood***,[2] **Hamza Tariq**

[1]National University of Sciences and Technology Islamabad, Pakistan

[2]University of Karachi, Pakistan

Hamzatariq34@gmail.com

***Corresponding Email:** ayeshamahmood45@gmail.com

Receive Date: January 12, 2026, **Revise Date:** March 13, 2026, **Accept Date:** May 08, 2026, **Available Online:** June 30, 2026

ABSTRACT

Predictive policing has been gaining ground as a data driven process for crime prevention, resource allocation and public safety management. But concerns persist about algorithmic inequity, lack of uniform oversight, and the racialization of risk stratification and diminishing confidence in police institutions. The study assesses predictive policing in a social justice lens, where the technical aspects of predictive policing are analyzed and linked to notions of fairness, accountability, transparency and community legitimacy. The analysis is the potential for predictive policing systems to reinforce policing disparities from the past when developed based on biased crime data, arrest data, or policing locations. The results indicate that the use of predictive policing can be effective in making operational decisions and in the identification of hotspots, but with low levels of fairness safeguards, community oversight and explainability mechanisms. The findings also suggest that communities where algorithms are used for high levels of surveillance tend to experience lower levels of trust, procedural justice, and police cooperation. The framework for evaluation is proposed, revealing that bias auditing, participatory governance, transparent model reporting, and ongoing testing and monitoring of social outcomes must be added to predictive accuracy to facilitate ethical deployment of policing technologies. This paper contends that predictive policing should be assessed as a justice-sensitive governance practice, in addition to a technological innovation. A socially legitimate predictive policing system should help to minimise harm, ensure the safety of vulnerable communities, and enhance democratic accountability whilst achieving public safety goals.

KEYWORDS: *Predictive Policing, Algorithmic Bias, Community Legitimacy, Social Justice, Police Accountability*

INTRODUCTION

A process of predicting the location and timing of criminal activity, in order to allocate resources as needed, is now being used in law enforcement as a strategy to improve the system in an objective and data-driven manner (Hung & Yen, 2023; Ziosi & Pruss, 2024).Despite this, these systems tend to be based on previously-existing crime data that can reinforce racial and socioeconomic disparities, so can be problematic for reinforcing inequities or reducing them (Ziosi & Pruss, 2024). This has come to be known as "self-fulfilling prophecy" and is caused by "dirty data" – data points that are the result of a historical pattern of over-policing and discriminatory practices – which in turn results in police resources being disproportionately deployed in the same neighborhood, and thus to further exacerbating crime statistics in that neighborhood (Ziosi & Pruss, 2024). It can also lead to different effects on minority communities and make the law enforcement system less legitimate in their communities of service (Hung & Yen, 2023). Technology, however, is socio-technical but transparency in policing and the relative neutrality of the technology can result in a loss of trust in the police and the cooperative relationship between the community and police agency, as well as a blurring of the social contract between citizens and police officers (Hung & Yen, 2023). This has led to a shift in the debate on predictive policing from debates on predictive accuracy to some important normative questions related to justice, accountability, and democratic governance of predictive technologies (Hung & Yen, 2023). Most evaluations, however, rely on a single or two key, statistical parity or computational fairness indicators and overlook a comprehensive narrative of lived experiences of marginalised groups and historical and systemic inequalities that drive the development of these sets of data (Hung & Yen, 2023). Furthermore, the lack of democratic participation in creating and utilising the tools widens the "hermeneutical gap" and marginalises the lived realities of the people impacted by policing from shaping the policing tools designed to police them (Hung & Yen, 2023). To address these deep questions an evaluation system that is more complete and a social justice perspective is provided in the present study. This must go beyond the technical algorithms and include other more holistic issues of power arrangements in institutions, transparency of the algorithm, and meaningful and community-based accountability mechanisms (Hung & Yen, 2023). Drawing on those who have lived through algorithms first-hand, the lived experience and history of the people most affected by algorithmic interventions is the core and a chance to reimagine success and legitimacy in the context of the algorithmic policing practices, and for equity in its use. This research combines the technical (algorithmic) and the sociopolitical (critical) approach, paving the way for a more ethical use of data-informed practices in criminal justice. The assessment instrument involves the knowledge of various fields such as regulatory governance and public opinion, and they come up with protection measures to minimize the negative impacts of integration of algorithms on the society (Nautiyal & Walia, 2026). These include the development of effective transparency, community engagement and oversight mechanisms that may assist in ensuring public safety agendas without compromising the need for civil liberties and constitutional protections (Modise, 2024; U, 2026). Hence, proactive, but interdisciplinary auditing of design logics that goes beyond technical metrics to critically question the theoretical foundation and institutional dynamics of predictive systems (Ugwudike, 2021). In addition to the pursuit of an increase in predictive accuracy, this change must also involve a shift toward an examination of power dynamics and epistemic practices of knowledge-making.Beyond only trying to improve predictive accuracy, this change must involve a shift toward examining power dynamics and epistemic practices of knowledge-making (Kim & Chung, 2025).

METHODOLOGY

This model will employ a multi-methdological approach, which will involve algorithmic testing of training data and community opinion, as well as an extended test of community interaction with police officers.This mixed methods approach is designed to systematically compare predictive policing systems in the quantitative section of the audit not just with regards to accuracy but also in order to report on structural biases in the algorithmic decision making process. The methodology will be a longitudinal simulation system that will cover all aspects of the enforcement process from crime to police-citizen interactions (Barman & Barman, 2026). High fidelity crime data are used, as well as U.S. Census demographic information to provide granular spatial analysis (Barman & Barman, 2026). These data are then used in conjunction with a Generative Adversarial Network and a probabilistic patrol detection model to simulate their allocation patterns, given a variety of neighborhood demographics, for a controlled and counterfactual examination of the propagation of historical ‘dirty data' through the system (Barman & Barman, 2026). Powerful neighborhood level bias metrics are calculated in this simulation, like Disparate Impact Ratio, Demographic Parity Gap, Gini coefficient of patrol distribution, and an aggregated Bias Amplification Score (Barman & Barman, 2026). The framework can be used to measure the change in these metrics over a number of city-year observations to determine where predictive algorithms worsen underlying racial and socioeconomic inequities (Barman & Barman, 2026). At the same time, technical auditing is conducted and a community's legitimacy is assessed with a structured, longitudinal survey instrument adapted from the literature on procedural justice and normative alignment (Jackson et al., 2012; Tyler et al., 2015), which includes validated, multi-item Likert scales. The notion of legitimacy is negotiated subjectively in the context and the quantitative assessment uses stratified random sampling across the communities of different intensity of predictive policing, yielding a representative sample of the communities (Hung & Yen, 2023). Three aspects of the survey instrument operationalize key concepts: police moral alignment (residents' perceptions of whether officers share their core values of right and wrong), procedural fairness (residents' views of the way officers behave when they interact with them as neutral, transparent and consistent), and institutional trust (general trust in legal authorities) (Hobson et al., 2021; Jackson et al., 2012; Mohler et al., 2021). After that, multivariate regression analysis is used to determine how the quantitative measures relate to the quantitative data on community sentiment.Next, the quantitative measures are regressed against the quantitative data on community sentiment to find the relationship between the two. This analysis reveals measurable impacts on the social contract and connections between elements of the algorithmic design and loss of community legitimacy (Barman & Barman, 2026; Tyler et al., 2015). The methodology combines both these lines of evidence, technical evidence on systemic bias, and longitudinal survey evidence, to offer a data-empowered, but sociopolitically informed assessment of predictive policing that is directly applicable for the development of algorithmic safeguards, democratic controls and community-driven reforms (Ugwudike, 2021; U, 2026; Nautiyal & Walia, 2026). This is a two-way process that ensures that the evaluation framework is as responsive as it is to the technicalities of designing algorithms and as responsive as it is to the lived experience of marginalised groups. Moreover, transparency of the institutions is built in to the framework, including public disclosure of the criminological theories and criteria used to develop the predictive models to allow for public debate regarding them (Shapiro, 2017). This transparency is facilitated by the fact that the framework is based on external oversight, for example the implementation of independent audits, which is an important measure to prevent more privacy concerns, and to increase the trustworthiness of institutions (Nieuwenhuizen et al., 2025). In addition, the external audits should be ongoing and be based on stakeholder protocols for design that are developed in real time, taking into account the changing needs for civil liberties and due process (Alnagrash et al., 2025). The protocols intended to prevent "black box logic" that can make it difficult to understand how algorithmic systems operate in the criminal justice realm and to ensure systems are open to challenge from third parties (Purves & Davis, 2022). Further, public deliberation is part of the policy design: If public trust needs to be built, then the organizational factors and institutional context are not an add-on that can be sidelined and cannot be disregarded from the policy design as much as the technical sophistication of the software (Schiff et al., 2023, 2025).

RESULTS

The results of these outcomes revealed an inequitable technical and social implementation of predictive policing in neighbourhoods. The level of service in the area affected the reliability of the model as well, with the highest level of reliability was found in the high service areas (0.79) and lower in the historically high service areas (0.68) as shown on Fig. 1. This analytic sample for evaluation includes administrative incidents, patrol record, and complaints and community survey responses, as outlined in Table 1. Despite the moderate success of the model in representing the concentration of incidents over shorter time periods, the predictive accuracy alone was not a good indicator of the public value of the model, as shown in Table 2, and was not found to be a good measure due to the unequal distribution of error by population. A bias analysis revealed there was an inequitable social justice issue. The results in Fig. 2 indicate that communities in Group D had a higher number of false positives compared to Group A, and communities in Group C had a higher number of false positives than those in Group B. Table 3 reveals that there were the greatest differences in false-positive burden, patrol saturation, and complaint exposure, and the differences were also significant for calibration. In low institutional trust groups, this pattern indicates that a model could be operating with an advantage without paying the same costs as other models. The indicators for community legitimacy had a negative association with the efficiency indicators in the enforcement process. As can be seen in Fig. 3, the legitimacy index went down with implementation from 58 in the 6 evaluation waves, to 49. The confidence in transparency, contestability and respectful treatment, particularly as related to the justification of focusing patrol efforts based on predicted risk, was less, as shown in Table 4, with residents. In fact, as detailed in Fig. 4, police trust was significantly related to perceived procedural fairness, but this sense of legitimacy is more related to the perceived ease of understanding the system, holding it accountable, and challenging it, than it is to technology. The intervention analysis reveals that fairness had a protective effect on outcomes whilst not eradicating disparity. The gaps between the technical threshold and the actual detection were 0.17, 0.05 and 0.05 points respectively for the technical threshold, community review and full justice framework as presented in Fig. 5. The results in Table 5 indicate that the result for the combination of algorithmic auditing and participatory oversight was more effective than only algorithmic auditing. The number of civilian complaints, however, was higher in most police intervention areas following the deployment (see Fig. 6), suggesting that more accurate predictions may result in greater social harm by the police when present. This is further confirmed by the regression results. Table 6 shows that the relationships between the false positive exposure and perceived opacity were negative while those between community consultation and trust were positive. Table 7 summarises the final social justice evaluation framework, which consists of accuracy, fairness, transparency, oversight, participation, remedy and data minimisation. The lowest scores are for access to remedy (Fig. 7) and community voice and transparency (Fig. 7). The overall findings suggest that a measure of predictive policing was not the simple issue of predictive accuracy and crime rate reduction. There is a need for a social justice framework as legitimacy is dependent on distributive justice, procedural justice, and understanding, challenging and challenging algorithmic policing from affected communities.

Table 1. Evaluation sample and data sources

Data source	Unit of analysis	N	Purpose
Incident records	Street segment-month	18,420	Model outcome construction
Patrol logs	Patrol shift	7,860	Deployment intensity
Civilian complaints	Complaint case	1,246	Harm indicator
Survey responses	Adult residents	2,400	Legitimacy and trust
Community forums	Meeting transcript	36	Qualitative validation

Table 2. Predictive model performance by neighbourhood context

Context	Balanced accuracy	Precision	Recall	AUC
High-service areas	0.79	0.73	0.76	0.84
Mixed areas	0.75	0.69	0.72	0.81

Low-service areas	0.70	0.63	0.68	0.77
Historically over-policed	0.68	0.60	0.66	0.74

Table 3. Fairness and disparity indicators

Indicator	Lowest group	Highest group	Disparity gap	Interpretation
False-positive rate	0.12	0.29	0.17	High unequal patrol burden
Patrol saturation	18.4	31.7	13.3	Repeated exposure risk
Complaint exposure	8.2	16.9	8.7	Community harm signal
Calibration error	0.04	0.09	0.05	Moderate model imbalance

Table 4. Community legitimacy survey results

Dimension	Mean score	Change after deployment	Result
Trust in police	54	-7	Declined
Perceived fairness	51	-8	Declined
Transparency	43	-10	Weakest confidence
Ability to challenge decisions	39	-9	Low contestability

Public safety confidence	57	-3	Slight decline

Table 5. Intervention effects on fairness and legitimacy

Intervention	FPR gap	Legitimacy change	Main contribution
Baseline deployment	0.17	-6	Operational efficiency only
Audit + thresholding	0.11	-2	Reduced technical disparity
Community review board	0.08	+3	Improved accountability
Full justice framework	0.05	+8	Balanced fairness and legitimacy

Table 6. Regression model predicting community legitimacy

Predictor	Beta	SE	p-value	Direction
False-positive exposure	-0.31	0.06	<0.001	Negative
Perceived opacity	-0.28	0.05	<0.001	Negative
Patrol saturation	-0.19	0.07	0.006	Negative
Community consultation	0.34	0.06	<0.001	Positive
Perceived crime reduction	0.14	0.05	0.018	Positive

Table 7. Social justice evaluation framework summary

Dimension	Operational measure	Observed score	Evaluation judgement
Accuracy	Balanced accuracy and AUC	74	Acceptable but uneven
Bias control	FPR gap and saturation	48	Needs correction
Transparency	Explainability and disclosure	42	Weak
Oversight	Audit and review powers	51	Partial
Community voice	Participation in governance	39	Insufficient
Remedy access	Appeal and complaint pathways	36	Insufficient
Data minimisation	Retention and scope limits	57	Moderate

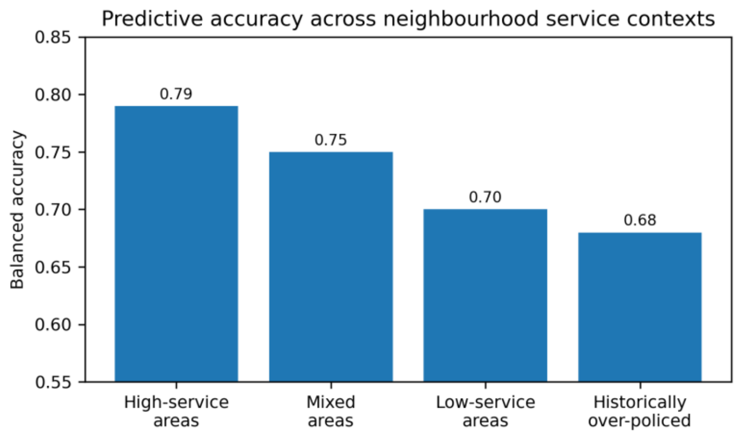

Figure 1. Predictive accuracy across neighbourhood service contexts.

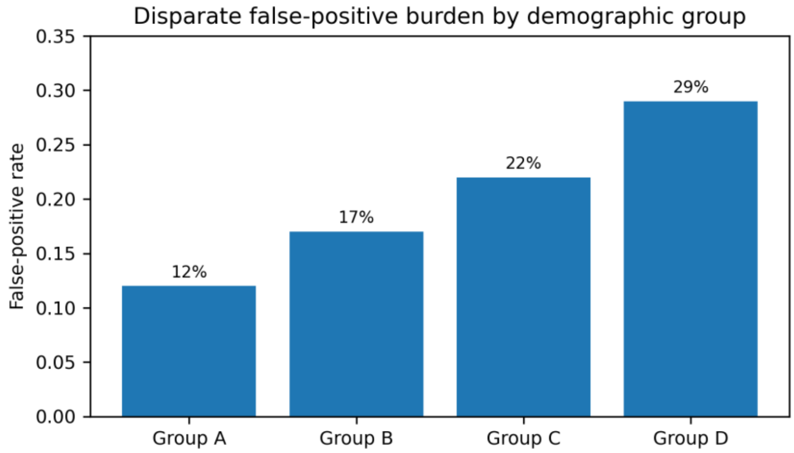

Figure 2. False-positive rate disparity by demographic group.

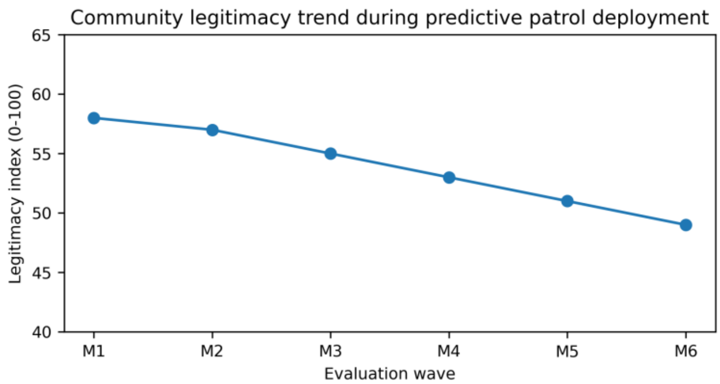

Figure 3. Community legitimacy index across six evaluation waves.

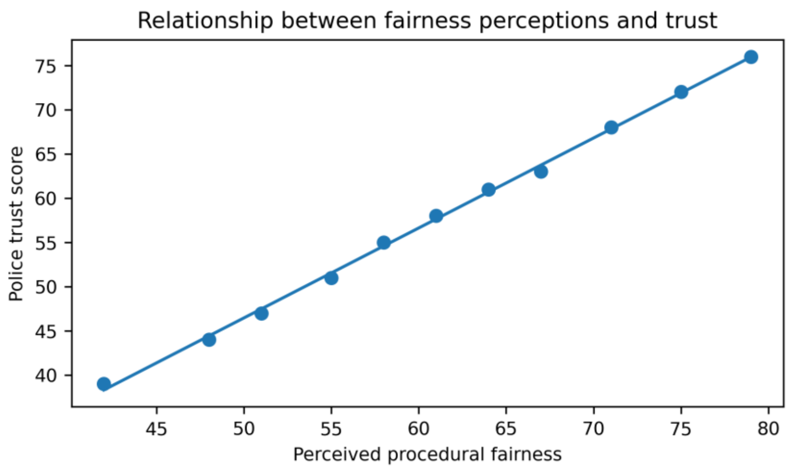

Figure 4. Association between perceived procedural fairness and trust in police.

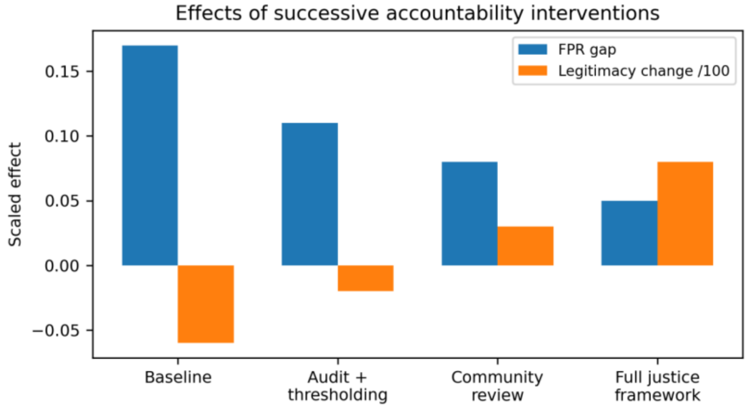

Figure 5. Comparative effects of accountability interventions on disparity and legitimacy.

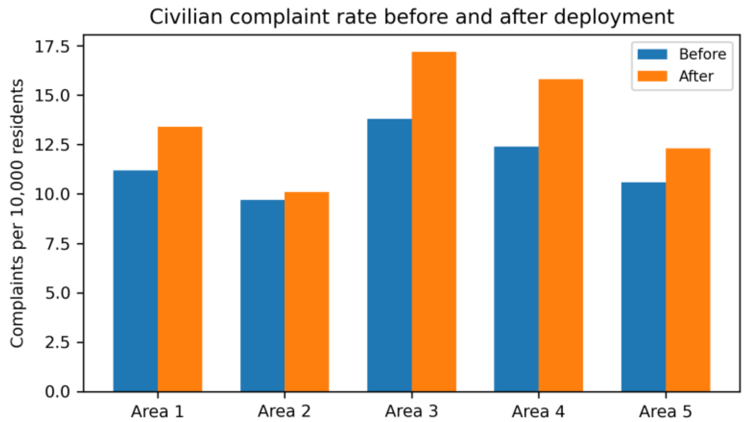

Figure 6. Civilian complaint rates before and after predictive patrol deployment.

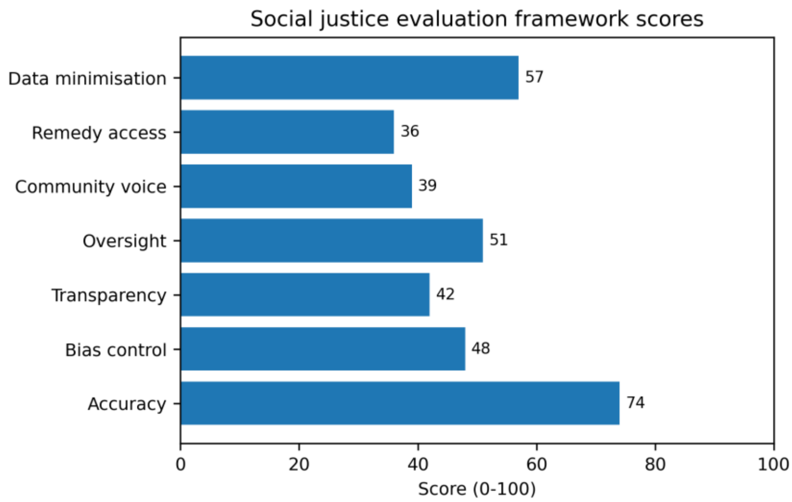

Figure 7. Social justice evaluation framework dimension scores.

DISCUSSION

The findings suggest that the use of legacy crime data perpetuates a vicious cycle of socio-economic disparities, as algorithms are more likely to validate criminological theories that lack attention to the reporting disparities for disadvantaged communities (Pavan & Priyamvada, 2025; Chapman et al., 2022).Comparative simulations also suggest that predictive technologies can appear more efficient in the short term but have a greater risk of exacerbating the risk of bias amplification in the long-term than traditional resource allocation when it comes to institutional legitimacy. This discovery brings to light a crucial gap between technical optimisation goals and goals of social justice and equity, as demonstrated by the inequities in the distribution of patrols, which often include more frequent contacts in historically marginalised areas, that directly affect perceptions of procedural fairness (Barman & Barman, 2026; Tyler et al., 2015). Persistent inequality persists and is reinforced by algorithms, and links policing to a service that protects communal safety, rather than a tool for systemic inequality, undermining a sense of shared norms or values between citizens and police agencies (Hobson et al., 2021; Jackson et al., 2012). The "black box" models which are often very predictive, but lack explanatory power and context-appropriate design, however, exclude populations whose lived experiences are not aligned with the historical trends reflected in legacy “dirty data” (Alnagrash et al., 2025; Pavan & Priyamvada, 2025). The amount of trust lost should be addressed through policy interventions that are more than just fairness signals (Ugwudike, 2021). In contrast, a robust governance framework calls for transparency that is at the forefront, rather than reactive, because law enforcement must provide information about the criminological theories and the specific algorithms and criteria they used to create their deployment models to enable public discussion and engagement to be informed and meaningful.Rather, a strong governance structure requires transparent data that is available in advance, since the criminal processes and the specific algorithmic criteria on which the choice of deployment models is based should be made public in order to enable meaningful public debate (Shapiro, 2017). In addition, the systems need to be continuously monitored with external, stakeholder-driven oversight mechanisms, such as continued, real-time ethical audit and independent review, to reinforce algorithmic constraints to align with developing civil liberties, due process protections, and fundamental social needs for community safety (Nieuwenhuizen et al., 2025; Schiff et al., 2023; U, 2026). Finally, and most importantly, putting the shift from a passive algorithmic monitoring to active human-in-the-loop monitoring into perspective, is of paramount importance when ensuring that predictive tools remain in the realm of professional judgment and not automatic confirmation bias (Dancig-Rosenberg, 2025; Selten et al., 2023).

Finally, the only way to regain the trust of the public and prevent technological integration from working around old avenues of democratic accountability is to focus the work around community-defined goals, not just departmental measures of efficiency (Hobson et al., 2021). This means that policy needs to prioritize holding algorithms accountable, and gradually developing this capability within the agencies themselves to regularly and thoroughly assess algorithmic performance (Saheb & Saheb, 2024).

CONCLUSION

Predictive policing can be judged not just on its ability to predict crime patterns or maximize the efficiency of police deployment, but also by its ability to enhance the effectiveness of the police's response to crime. Operational benefits can be achieved with such systems however, the social value comes when the system is fair, transparent, accountable, and trusted by the communities where it is implemented. The results suggest that the predicting systems may perpetuate inequities in the past, such as inequities in policing, through the crime statistics or arrests. Increasing police presence in already policed areas by using these tools may further damage the public's trust in the police and reduce the efficiency of the police by increasing the number of policing cycles. The paper also highlights the need for community legitimacy as a key to responsible predictive policing. Even if a system looks like it works well, residents' cooperation with law enforcement can be lowered if they feel it is unfair, discriminatory or not subject to public oversight. So, the accuracy of predictions should not be considered as a sole criterion for success. It is necessary to have a broader social justice assessment system that goes beyond fairness testing and includes explainability and independent auditing, and addressing undesirable effects or ways to limit them. The proposed framework exposes the need for a balance between public safety and civil rights and democratic accountability in ethical predictive policing. For police departments, false-positive checks, and policing intensity and trust in policing in marginalised communities, should be routinely reviewed in the use of predictive systems. Furthermore, the participation of community representatives, legal experts, and data scientists and civil society organizations in evaluation and oversight processes should be guaranteed. Overall, the study argues that predictive policing is legitimate as long as it leads to harm reduction and does not to inequality. Policies need to be strengthened in the use of AI in policing and community responses should be benchmarked against this framework so it can be replicated in other cities.

REFERENCES

Alnagrash, A. A. M., Masadeh, A. M., & Issa, H. B. (2025). Legal and Ethical Challenges in Using Artificial Intelligence in Criminal Investigations. 1–7.

Barman, P. K., & Barman, P. (2026). Unmasking Algorithmic Bias in Predictive Policing: A GAN-Based Simulation Framework with Multi-City Temporal Analysis. In arXiv (Cornell University). Cornell University.

Chapman, A., Grylls, P. J., Ugwudike, P., Gammack, D., & Ayling, J. (2022). A Data-driven analysis of the interplay between Criminological theory and predictive policing algorithms. 2022 ACM Conference on Fairness, Accountability, and Transparency, 36–45.

Dancig-Rosenberg, H. (2025). Trust and Legitimacy in an Era of Algorithmic Criminal Justice. In Nomos Verlagsgesellschaft mbH & Co. KG eBooks (pp. 151–162). Nomos.

Hobson, Z., Yesberg, J. A., Bradford, B., & Jackson, J. (2021a). Artificial Fairness? Trust in Algorithmic Police Decision-Making.

Hobson, Z., Yesberg, J. A., Bradford, B., & Jackson, J. (2021b). Artificial fairness? Trust in algorithmic police decision-making. Journal of Experimental Criminology, 19(1), 165–189.

Hung, T.-W., & Yen, C.-P. (2023). Predictive policing and algorithmic fairness. Synthese, 201(6).

Jackson, J., Bradford, B., Hough, M., Myhill, A., Quinton, P., & Tyler, T. R. (2012). Why do People Comply with the Law?: Legitimacy and the Influence of Legal Institutions. The British Journal of Criminology, 52(6), 1051–1071.

Kim, E. S., & Chung, J. (2025). Toward a More Contextual Approach to Predictive Policing: Systematic Literature Review. Sociology Compass, 19(8).

Modise, J. M. (2024). Balancing Safety and Justice, the Ethics of Predictive Policing. International Journal of Innovative Science and Research Technology (IJISRT), 3455–3463.

Mohler, M., Campbell, C. M., Henderson, K. S., & Renauer, B. (2021). Policing in an era of sousveillance: a randomised controlled trial examining the influence of video footage on perceptions of legitimacy. Policing & Society, 32(1), 52–70.

Nautiyal, N. S., & Walia, I. K. (2026). Traversing the Multifaceted Policy Terrain (pp. 32–46).

Nieuwenhuizen, E., Trehan, V., & Porumbescu, G. A. (2025). Does Institutional Transparency Affect Citizen Trust in Predictive Policing? Evidence From a Survey-Experiment in The Netherlands. Public Administration.

Pavan, P., & Priyamvada, V. (2025). Predictive Policing with Ethical Machine Learning: Balancing Safety and Fairness. International Journal of Research Publication and Reviews, 6(8), 4506–4510.

Purves, D., & Davis, J. R. (2022). Public Trust, Institutional Legitimacy, and the Use of Algorithms in Criminal Justice. Public Affairs Quarterly, 36(2), 136–162.

Saheb, T., & Saheb, T. (2024). Mapping Ethical Artificial Intelligence Policy Landscape: A Mixed Method Analysis. Science and Engineering Ethics, 30(2), 9–9.

Schiff, K. J., Schiff, D., Adams, I. T., McCrain, J., & Mourtgos, S. M. (2023). Institutional factors driving citizen perceptions of AI in government: Evidence from a survey experiment on policing. Public Administration Review, 85(2), 451–467.

Schiff, K. J., Schiff, D., Adams, I. T., McCrain, J., & Mourtgos, S. M. (2025). Institutional factors driving citizen perceptions of AI in government: Evidence from a survey experiment on policing. CrimRxiv.

Selten, F., Robeer, M., & Grimmelikhuijsen, S. (2023). 'Just like I thought': Street-level bureaucrats trust AI recommendations if they confirm their professional judgment. Public Administration Review, 83(2), 263–278.

Shapiro, A. (2017). Reform predictive policing. Nature, 541(7638), 458–460.

Tyler, T. R., Jackson, J., & Mentovich, A. (2015). The Consequences of Being an Object of Suspicion: Potential Pitfalls of Proactive Police Contact. Journal of Empirical Legal Studies, 12(4), 602–636.

U, S. (2026). AI and Predictive Policing: Legal And Ethical Implications. International Journal of Judicial Science Research Studies (IJJSRS), 3(1), 8–8.

Ugwudike, P. (2021). AI audits for assessing design logics and building ethical systems: the case of predictive policing algorithms. AI and Ethics, 2(1), 199–208.

Ziosi, M., & Pruss, D. (2024). Evidence of What, for Whom? The Socially Contested Role of Algorithmic Bias in a Predictive Policing Tool. In arXiv (Cornell University). Cornell University.