

Applications of Data Mining and Machine Learning in State Surveillance and the Oppression of Marginalized Populations: A Systematic Literature Review

Abstract

The expansion of data mining and machine learning (ML) technologies has fundamentally transformed state surveillance capabilities. Governments increasingly deploy predictive analytics, biometric identification, social network analysis, and natural language processing to monitor populations, manage risk, and enforce social control. While these technologies promise efficiency and security, a growing body of scholarship argues that they disproportionately target and harm marginalized populations. This systematic literature review synthesizes interdisciplinary research examining how data mining and ML are applied in state surveillance systems and how these applications contribute to structural inequality and repression. Following PRISMA-inspired review standards, this study identifies dominant applications, analyzes mechanisms of harm, evaluates methodological trends, and highlights research gaps. The findings demonstrate that predictive policing, algorithmic welfare surveillance, border analytics, and biometric governance are central domains in which ML systems amplify historical biases, institutionalize discrimination, and reduce procedural transparency. The review concludes with a testable hypothesis concerning the relationship between predictive policing intensity and racialized arrest disparities, offering a foundation for empirical investigation.

1. Introduction

The integration of machine learning and large-scale data mining into governance has reconfigured the landscape of state surveillance. Traditionally, surveillance relied on human intelligence gathering and bureaucratic recordkeeping. Today, digital infrastructures—mass data collection, real-time analytics, automated risk scoring—enable states to process vast quantities of behavioral data at unprecedented scale and speed.

Scholars across sociology, computer science, law, and political theory argue that these technological transformations are not neutral. Rather, they often embed historical inequalities into automated systems, intensifying scrutiny and coercive intervention in communities already subject to structural marginalization. Marginalized populations—particularly racial minorities, migrants, low-income individuals, religious minorities, and political dissidents—are disproportionately affected by algorithmic decision-making in policing, welfare administration, border control, and public security.

This review addresses five research questions concerning applications, impacts, methodological trends, research gaps, and future empirical testing.

2. Methodology

This study follows systematic review principles inspired by PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), adapted for interdisciplinary social science research.

Searches were conducted across major academic databases including Scopus, Web of Science, JSTOR, IEEE Xplore, ACM Digital Library, and Google Scholar. Keywords combined terms such as “machine learning,” “state surveillance,” “predictive policing,” “algorithmic governance,” “facial recognition,” and “welfare surveillance.”

Inclusion criteria required peer-reviewed or high-impact policy sources (2005–2024), explicit discussion of marginalized populations, and analytical rigor. Exclusion criteria removed purely technical optimization studies and non-state surveillance contexts.

Initial searches yielded approximately 1,240 records. After screening and full-text review, 126 sources were retained for thematic coding and qualitative synthesis.

3. Key Application Domains

Major domains identified include predictive policing and criminal justice risk assessment, biometric and facial recognition surveillance, welfare fraud detection and eligibility automation, border analytics and migration control, and political social media monitoring.

Across domains, studies consistently document patterns of bias reproduction, feedback loops, disproportionate error rates affecting marginalized groups, and limited avenues for procedural redress.

4. Cross-Cutting Mechanisms of Oppression

The literature identifies five recurring mechanisms: (1) historical data bias encoding structural discrimination; (2) feedback loops reinforcing targeted policing and monitoring; (3) algorithmic opacity limiting accountability; (4) unequal distribution of errors; and (5) normalization of suspicion toward marginalized populations.

These mechanisms collectively institutionalize inequality through automated governance systems.

5. Methodological Trends and Gaps

Research methods include quantitative bias audits, ethnographic fieldwork, legal analysis, and critical sociotechnical theory. However, significant gaps remain.

These include underrepresentation of Global South contexts, limited intersectional analysis, scarcity of longitudinal causal studies, and insufficient participatory research involving affected communities.

6. Proposed Testable Hypothesis

H1: Jurisdictions that implement predictive policing systems will experience a statistically significant increase in racial disparities in low-level arrest rates within five years of implementation, compared to demographically similar jurisdictions without predictive policing.

Operationalization would involve a difference-in-differences regression model using panel data, controlling for crime rates, socioeconomic conditions, policing budgets, and demographic composition.

7. Conclusion

This systematic review demonstrates that machine learning and data mining technologies are deeply embedded in contemporary state surveillance infrastructures. Evidence across domains suggests these systems frequently reproduce and intensify structural inequalities affecting marginalized populations.

Future research should prioritize longitudinal empirical testing, intersectional frameworks, comparative policy analysis, and community-centered methodologies to better assess and mitigate harms associated with algorithmic state surveillance.