Auditing the Long-Term Societal Impact of AI-Driven Surveillance through Temporal Bias Analysis and Human-in-the-Loop Governance

Theodore Pennington
Department of Computer Science and Artificial Intelligence
University of New Mexico
theodore.p@unm.edu

Abstract

The proliferation of artificial intelligence within global surveillance infrastructures has transitioned from localized security enhancements to pervasive socio-technical systems that govern public life. While much of the existing discourse focuses on immediate algorithmic fairness and data privacy, there is a critical need to examine the long-term societal impacts through the lens of temporal bias and the erosion of institutional accountability. This paper investigates the systemic risks associated with the continuous deployment of automated monitoring systems, arguing that bias is not merely a static artifact of training data but a dynamic phenomenon that evolves over time through feedback loops between algorithmic outputs and human behavioral shifts. We propose a comprehensive auditing framework centered on temporal bias analysis—measuring how surveillance accuracy and societal stratification fluctuate across extended horizons. Furthermore, the research advocates for a robust Human-in-the-Loop (HITL) governance model that moves beyond symbolic oversight toward a functional integration of human judgment at critical decision-making junctions. By analyzing the infrastructure and policy implications of these systems, the paper explores how the intersection of automated decision-making and public space surveillance challenges traditional notions of civil liberty and institutional robustness. We conclude that without a structural shift toward temporal auditing and active human intervention, the long-term deployment of AI surveillance risks entrenching historical inequities and creating rigid, non-adaptive governance structures that are ill-equipped to handle the complexities of evolving social norms.

Keywords:

AI Surveillance, Temporal Bias, Human-in-the-Loop, Algorithmic Governance, Socio-Technical Infrastructure, Fairness and Accountability.

1. Introduction

The integration of artificial intelligence into the fabric of urban and digital surveillance represents one of the most significant shifts in the history of governance and public safety [21]. What began as a tool for automating mundane monitoring tasks has evolved into a

complex ecosystem of predictive analytics, biometric identification, and behavioral forecasting [8]. These systems do not operate in a vacuum; they are embedded within existing social hierarchies and physical infrastructures, creating a dual-layered reality where digital logic increasingly dictates physical movement and access [19]. The central challenge of this era is not simply the existence of surveillance, but the opaque nature of the AI agents that manage it and the enduring, often invisible, impacts they have on the social contract over decades [24].

Current academic inquiry often prioritizes the technical optimization of these systems—increasing accuracy, reducing latency, or improving object detection—while neglecting the longitudinal effects of these technologies on the societies they inhabit [18]. This research addresses this gap by focusing on the concept of temporal bias. Unlike cross-sectional bias, which examines a snapshot of a system's performance at a single point in time, temporal bias looks at how algorithmic performance degrades or skews as social environments change [3]. As AI-driven surveillance systems interact with the public, they alter the very behaviors they were designed to monitor, creating a recursive loop that can amplify initial inaccuracies into systemic failures [12].

Furthermore, the governance of these systems has largely remained reactive [31]. Regulatory frameworks often struggle to keep pace with the rapid iteration of large-scale models and the decentralized nature of their deployment [32]. This paper argues that the traditional approach to AI auditing is insufficient for systems that possess such high societal stakes [25]. Instead, we propose a governance architecture that centers on Human-in-the-Loop (HITL) principles, ensuring that human agency is not merely a peripheral check but a fundamental component of the system’s adaptive logic [2]. Through a detailed analysis of engineering trade-offs and policy implications, this paper seeks to provide a roadmap for more resilient and equitable surveillance infrastructures.

2. The Socio-Technical Infrastructure of AI Surveillance

To understand the long-term impact of AI-driven surveillance, one must first view it as a socio-technical infrastructure rather than a mere collection of software tools [29]. This infrastructure includes the physical hardware of sensors, the data pipelines that transport information to centralized or edge-computing nodes, and the institutional protocols that determine how resulting insights are used by law enforcement or private entities [5]. The sustainability of such a system depends on its ability to maintain public trust and operational integrity over long durations [17]. However, the opaque nature of many deep learning models creates a fundamental tension between the need for security and the requirement for institutional transparency [24].

The architectural complexity of these systems often masks the power dynamics at play [28]. When surveillance is automated, the "gaze" of the state or the corporation becomes ubiquitous and tireless [12]. This ubiquity leads to a phenomenon known as chilling effects, where individuals self-censor their behavior in public spaces due to the awareness of constant monitoring [21]. Over time, this alters the cultural and political vibrancy of urban

environments, as the threshold for what is considered normal behavior is narrowed by the algorithmic definitions of deviance [14]. The infrastructure is not neutral; it encodes the values and priorities of its designers, which may be at odds with the evolving needs of a diverse populace [23].

Robustness in this context refers not only to the technical uptime of the system but also to its social robustness—its ability to function without causing systemic harm or triggering widespread social backlash [16]. Engineering trade-offs frequently favor high-precision monitoring over privacy-preserving architectures, under the assumption that more data inherently leads to more safety [8]. Yet, the history of large-scale systems suggests that data saturation can lead to noise that overwhelms human oversight, leading to a reliance on automated alerts that may be plagued by systemic bias [11]. The transition toward edge-computing further complicates these dynamics, as governance becomes more difficult to centralize and audit [19].

3. Temporal Bias: The Dynamics of Algorithmic Decay

Temporal bias represents a critical, yet understudied, dimension of algorithmic fairness. Most contemporary audits focus on static metrics—asking whether a model performs equally well across different demographic groups at the time of testing [7]. However, social environments are dynamic; populations shift, fashion trends change, and social norms regarding behavior and movement evolve [4]. A surveillance model trained on data from one decade may become increasingly biased or inaccurate as it encounters a world that no longer matches its training distribution [31]. This concept drift is particularly dangerous in surveillance, where a decrease in accuracy can lead to false accusations or the over-policing of certain communities [11].

The long-term societal impact of this decay is profound. When a surveillance system's performance degrades, it does not do so uniformly [16]. Historically marginalized groups often bear the brunt of algorithmic errors, as the data used to train these systems frequently underrepresents these populations or overrepresents them in negative contexts [23]. As the system continues to operate, these errors are fed back into the training loops of future models, creating a self-reinforcing cycle of discrimination [11]. This is the essence of temporal bias: the gradual entrenchment of historical prejudices through the medium of supposedly objective technology [8].

Auditing for temporal bias requires a shift in methodology. It necessitates the creation of longitudinal datasets and the implementation of continuous monitoring protocols that track system performance against a moving baseline of social reality [26]. This involves not just technical metrics, but a qualitative understanding of how the presence of AI surveillance changes the data it collects [14]. For instance, if people avoid certain areas because they know cameras are equipped with facial recognition, the resulting data suggests those areas are safer, which may lead to a reduction in actual community services. Understanding these feedback loops is essential for any meaningful audit of the societal footprint of artificial intelligence [13].

4. Human-in-the-Loop (HITL) Governance Models

The limitations of purely automated surveillance necessitate a more sophisticated approach to governance: the Human-in-the-Loop (HITL) model [22]. This framework posits that for high-stakes AI systems, human judgment must be integrated at key intervention points to provide the nuance, empathy, and ethical reasoning that machines lack [18]. In the context of surveillance, this means that an AI should not be the final arbiter of identity, intent, or threat level. Instead, it should serve as a decision-support tool that provides evidence for human review [25].

However, the implementation of HITL is often hindered by automation bias, where human operators become overly reliant on algorithmic suggestions, treating them as infallible truths rather than probabilistic estimates [22]. To counter this, governance structures must be designed to promote active critical engagement [2]. This requires a rethink of the institutional workflows surrounding surveillance. Rather than having a human simply confirm an AI alert, a robust HITL system would require independent human verification based on raw data, using the AI output only as a secondary reference point [33]. This ensures that the human remains the primary agent of accountability [10].

Furthermore, HITL governance must extend beyond the operational level to the policy and design levels [20]. This involves diverse stakeholders—including community advocates, ethicists, and legal experts—in the initial development and ongoing auditing of surveillance systems [15]. By incorporating a wide range of human perspectives, the system can be better calibrated to respect civil liberties and cultural nuances [29]. This interdisciplinary approach to governance is essential for maintaining the legitimacy of surveillance infrastructures in a democratic society [32]. It also provides a mechanism for correcting the temporal biases identified earlier, as human oversight can recognize social shifts that a static model might miss [33].

5. Systemic Robustness and Deployment Challenges

The deployment of AI-driven surveillance at scale introduces unique challenges to systemic robustness [1]. Large-scale infrastructures are inherently fragile when they rely on highly centralized models that can be compromised or misconfigured [24]. In a surveillance context, robustness refers to the system's ability to resist adversarial attacks, handle sensor degradation, and maintain consistent performance across varying environmental conditions [30]. However, the pursuit of robustness often creates a technological lock-in, where institutions become so dependent on a specific vendor or architecture that they are unable to adapt to new ethical or legal requirements [17].

From an engineering perspective, the trade-off between sensitivity and specificity is constant. A system tuned to catch every possible threat will inevitably produce a high volume of false positives, leading to the harassment of innocent individuals [6]. Conversely, a system that is too conservative may miss genuine safety risks. The long-term societal cost of these trade-offs is rarely calculated [11]. Over decades, a high false-positive rate for a specific sub-population can lead to deep-seated distrust in public institutions and a breakdown of community relations

[4]. Deployment strategies must therefore prioritize social safety alongside technical efficacy, ensuring that the fail-safes of the system protect individual rights as much as they protect the physical infrastructure [25].

Moreover, the sustainability of AI surveillance is tied to its data management practices [19]. The storage and processing of massive amounts of biometric and behavioral data present significant security risks [28]. As these systems age, the legacy data they hold becomes a liability [10]. A robust governance framework must include data sunsetting policies—mandating the deletion of information after its immediate utility has passed—to prevent the creation of permanent, unalterable digital shadows for every citizen [21]. The long-term audit must examine not just the AI's current actions, but the enduring presence of the data it has already consumed [13].

6. Policy Implications and the Future of Civil Liberties

The intersection of AI surveillance and public policy is where the most significant societal impacts are decided [20]. Current legal frameworks are often ill-equipped to handle the nuances of algorithmic monitoring [32]. Most privacy laws were written for an era of physical documents or static digital records, not for an era of real-time, predictive behavioral analysis [10]. There is a pressing need for algorithmic impact assessments that are legally mandated before any large-scale surveillance system is deployed [29]. These assessments should be treated like environmental impact reports, requiring a thorough investigation of how the system will affect the social ecology of the area [16].

One of the most concerning policy implications is the potential for function creep, where a system designed for a specific purpose—such as traffic management—is gradually expanded to include criminal tracking or political monitoring [21]. Without clear, enforceable boundaries, the infrastructure of surveillance tends to expand until it covers all aspects of public life [12]. Temporal auditing is a vital tool for identifying and halting function creep, as it tracks the evolution of a system's use cases over time [26].

The future of civil liberties in an age of AI surveillance depends on the ability to maintain contextual integrity [7]. This principle suggests that privacy is not just about the secrecy of information, but about the appropriate flow of information within specific social contexts [18]. When an AI system crosses these boundaries—for instance, by using medical data to inform public surveillance—it violates the social contract [23]. Policy must therefore focus on creating hard boundaries between different data domains and ensuring that the human-in-the-loop has the legal authority to override algorithmic decisions that violate these norms [31].

7. Strategic Auditing and Longitudinal Analysis

A strategic audit of AI-driven surveillance must go beyond checking for bugs and instead evaluate the system's overall health within the social body [25]. This requires a multidisciplinary approach that combines data science, sociology, and legal theory [29]. The audit should be periodic and transparent, with the results made available to the public to foster

institutional trust [26]. Longitudinal analysis is the cornerstone of this process, as it allows researchers to see the slow-motion effects of surveillance that are invisible in short-term studies [4].

In conducting these audits, researchers must pay particular attention to the proxy variables that AI systems often use [3]. Even if a system is programmed to ignore protected characteristics, it may still develop biases based on variables that are highly correlated with demographic groups [23]. A long-term audit would track how these proxies evolve and whether the system is inadvertently creating new forms of digital redlining [11]. This requires a high level of technical sophistication and a deep commitment to social equity from the auditing body [27].

Furthermore, the audit must evaluate the exit strategy for these systems [15]. What happens when a surveillance technology is found to be harmful? The infrastructure should be designed with modularity in mind, allowing for the removal or replacement of biased components without the collapse of the entire security framework [17]. This engineering flexibility is a prerequisite for responsible governance [1]. The ultimate goal of the audit is to ensure that AI serves as a tool for collective well-being rather than an instrument of unaccountable control [11].

8. The Role of Path-Level Intervention in Safety

Recent advancements in AI safety have highlighted the importance of intervention at the path level rather than just the input or output level [30]. In complex surveillance systems, safety is often compromised not by a single erroneous data point, but by the cumulative weight of the path the data takes through the network [15]. By implementing robust safety protocols that intervene at the path-level, developers can provide a higher degree of assurance that the model will not reach harmful conclusions or amplify biases during its processing stages [20].

This approach aligns with the need for more granular control within surveillance infrastructures [24]. If a model's decision-making path is found to be relying on biased heuristics, a path-level intervention can redirect the logic toward more neutral ground before a final decision is reached [30]. This is a technical manifestation of the Human-in-the-Loop philosophy, where human-defined safety constraints are baked into the very architecture of the neural network [22]. As we move toward more autonomous surveillance agents, these types of robust safety measures will become the primary defense against unintended societal consequences [18].

The integration of such safety protocols also enhances the auditability of the system [25]. If an auditor can trace the path of a decision and identify exactly where a bias was introduced, the black box of AI becomes significantly more transparent [24]. This level of transparency is essential for legal accountability and for the public's ability to challenge algorithmic decisions [32]. In the long run, the robustness of our surveillance systems will be measured by our ability to intervene and correct their internal logic in real-time [1].

9. Engineering for Fairness and Sustainable Governance

Engineering for fairness is not a one-time task but a continuous process of calibration and refinement [3]. It requires a fundamental shift in the design philosophy of AI-driven surveillance [6]. Instead of optimizing solely for efficiency, engineers must optimize for equity [23]. This involves incorporating fairness constraints directly into the objective functions of the models and using diverse datasets that reflect the actual complexity of the human population [11].

Sustainability in governance also requires a sustainable workforce [13]. The human-in-the-loop must be well-trained, adequately compensated, and mentally supported to handle the rigors of monitoring sensitive data [22]. If the human operators are overworked or under-trained, the HITL model becomes a mere formality, and the system effectively reverts to full automation [13]. Therefore, the long-term impact of surveillance is also tied to the labor practices of the institutions that manage it [27].

Finally, the governance of AI surveillance must be global in scope [18]. As these technologies are exported across borders, the biases and values of the exporting nations are often encoded within them [21]. International standards for AI surveillance auditing are needed to prevent the spread of digital authoritarianism and to ensure that human rights are protected globally [20]. A sustainable future for AI is one where technology is a force for democratization and safety, grounded in the principles of transparency, accountability, and human-centric design [15].

10. Conclusion

The long-term societal impact of AI-driven surveillance is one of the defining challenges of the twenty-first century. As these systems become more deeply embedded in our infrastructures, the risks of temporal bias and the erosion of human agency grow exponentially. This paper has argued that a purely technical approach to AI safety is insufficient. Instead, we must embrace a socio-technical perspective that prioritizes longitudinal auditing, temporal bias analysis, and robust Human-in-the-Loop governance.

By viewing surveillance through the lens of large-scale systems and interdisciplinary research, we can identify the structural trade-offs that threaten the social contract. The path toward a more equitable and resilient future involves a commitment to transparency, a willingness to intervene in algorithmic processes, and a dedication to protecting civil liberties in a digital age. The human must remain at the center of the loop, not just as a supervisor, but as the final moral and ethical authority. Only then can we ensure that the power of AI is used to enhance the security and freedom of all members of society, rather than to entrench the inequities of the past.

References

1. Ajunwa, I. (2023). The Quantified Worker: Law and Technology in the Modern Workplace. Cambridge University Press.

2. Amershi, S., et al. (2019). Guidelines for Human-AI Interaction. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.

3. Barocas, S., Hardt, M., & Narayanan, A. (2024). Fairness and Machine Learning: Limitations and Opportunities. MIT Press.

4. Benjamin, R. (2019). Race After Technology: Abolitionist Tools for the New Jim Code. Polity.

5. Bowker, G. C., & Star, S. L. (2000). Sorting Things Out: Classification and Its Consequences. MIT Press.

6. Broussard, M. (2018). Artificial Unintelligence: How Computers Misunderstand the World. MIT Press.

7. Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Proceedings of Machine Learning Research.

8. Crawford, K. (2021). The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence. Yale University Press.

9. Dastin, J. (2018). Amazon scraps AI recruiting tool that showed bias against women. Reuters.

10. Edwards, L., & Veale, M. (2017). Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For. Duke Law & Technology Review.

11. Eubanks, V. (2018). Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. St. Martin's Press.

12. Gandy, O. H. (2021). The Panoptic Sort: A Political Economy of Personal Information. Oxford University Press.

13. Gray, M. L., & Suri, S. (2019). Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass. Houghton Mifflin Harcourt.

14. Green, B. (2022). The Smart Enough City: Putting Technology in Its Place to Reclaim Our Urban Future. MIT Press.

15. Hadfield-Menell, D., et al. (2016). The Cooperative Inverse Reinforcement Learning Problem. Advances in Neural Information Processing Systems.

16. Hoffmann, A. L. (2019). Where fairness fails: Data, algorithms, and the limits of

antidiscrimination discourse. Information, Communication & Society.

17. Jasanoff, S. (2016). The Ethics of Invention: Technology and the Human Future. W. W. Norton & Company.

18. Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence.

19. Kitchin, R. (2014). The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences. SAGE Publications.

20. Leslie, D. (2019). Understanding artificial intelligence ethics and safety. The Alan Turing Institute.

21. Lyon, D. (2018). The Culture of Surveillance: Watching as a Way of Life. Polity.

22. Mindell, D. A. (2015). Our Robots, Ourselves: Robotics and the Myths of Autonomy. Viking.

23. Noble, S. U. (2018). Algorithms of Oppression: How Search Engines Reinforce Racism. NYU Press.

24. Pasquale, F. (2015). The Black Box Society: The Secret Algorithms That Control Money and Information. Harvard University Press.

25. Raji, I. D., et al. (2020). Closing the AI accountability gap: Defining challenges for internal algorithmic auditing. Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency.

26. Sandvig, C., et al. (2014). Auditing Algorithms: Research Methods for Detecting Discrimination on Algorithmic Platforms. 64th Annual Meeting of the International Communication Association.

27. Selwyn, N. (2019). Should Robots Replace Teachers? Polity.

28. Shoshana, Z. (2019). The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power. PublicAffairs.

29. Selbst, A. D., et al. (2019). Fairness and Abstraction in Sociotechnical Systems. Proceedings of the 2019 Conference on Fairness, Accountability, and Transparency.

30. Shi, C., Li, S., Lu, W., Wu, W., Wang, C., Cheng, Z., ... & Chua, T. S. (2026). TraceRouter: Robust Safety for Large Foundation Models via Path-Level Intervention. arXiv preprint arXiv:2601.21900.

31. Tsamados, A., et al. (2022). The ethics of algorithms: Key problems and solutions. AI & Society.

32. Wachter, S., Mittelstadt, B., & Russell, C. (2021). Why Fairness Cannot Be Automated: Bridging the Gap Between EU Non-Discrimination Law and AI. Computer Law & Security Review.

33. Whittaker, M., et al. (2018). AI Now Report 2018. AI Now Institute.