AI IN PUBLIC HEALTH SURVEILLANCE: BALANCING INNOVATION, PRIVACY, AND FUNDAMENTAL RIGHTS

Shanu Singh Chouhan, National Law Institute University, Bhopal

ABSTRACT

Artificial intelligence (AI) has become a central instrument of contemporary public health surveillance, powering disease-outbreak detection, syndromic monitoring, predictive epidemic modelling, contact tracing, and resource allocation. The COVID-19 pandemic accelerated the adoption of AI-enabled surveillance tools worldwide, demonstrating both their capacity to detect and contain threats faster than traditional systems and their potential to intrude upon privacy, entrench discrimination, and erode fundamental rights. This paper examines the dual character of AI in public health surveillance, tracing its major applications, the privacy and human-rights risks these applications generate, and the evolving regulatory architecture-principally the World Health Organization's ethics guidance, the European Union's General Data Protection Regulation and Artificial Intelligence Act, and emerging national frameworks-designed to govern them. Drawing on peer-reviewed literature, international guidance documents, and case studies of contact-tracing applications and AI-based early-warning systems such as BlueDot and HealthMap, the paper argues that innovation and rights protection are not inherently opposed but require deliberate institutional design: privacy-enhancing technologies, proportionality-based legal tests, algorithmic bias audits, transparency mechanisms, and participatory governance. The paper concludes that a rights-respecting approach to AI-enabled surveillance is not merely a constraint on innovation but a precondition for its long-term legitimacy and effectiveness.

Keywords: artificial intelligence, public health surveillance, privacy, data protection, fundamental rights, GDPR, AI Act, contact tracing, algorithmic bias

1. INTRODUCTION

Public health surveillance-the systematic, ongoing collection, analysis, and interpretation of health data for the purpose of planning, implementing, and evaluating public health action-has always depended on the tension between two imperatives: the collective need to detect and respond to threats to population health, and the individual's interest in controlling information about their body, behaviour, and associations. For most of the twentieth century, this tension was managed through comparatively slow, human-mediated systems: notifiable disease reporting, sentinel clinics, and manual contact tracing. The twenty-first century has transformed this landscape. Artificial intelligence-understood broadly as computational systems capable of performing tasks that typically require human intelligence, including pattern recognition, prediction, and natural language processing-now underlies disease-forecasting models, automated early-warning platforms that scan global media in real time, machine-learning-based risk stratification tools, and mobile applications capable of reconstructing an individual's proximity history over weeks or months.[1]

The promise of these tools is considerable. AI systems such as BlueDot and HealthMap have demonstrated the capacity to detect emerging outbreaks days before official government or WHO announcements, as occurred with the identification of an unusual pneumonia cluster in Wuhan in December 2019.[2] During the COVID-19 pandemic, dozens of governments deployed AI-enabled contact-tracing and exposure-notification applications, syndromic surveillance dashboards, and predictive models to allocate scarce medical resources.[3] Machine learning has similarly been applied to dengue and influenza forecasting, misinformation detection, and vaccine-uptake prediction.

Yet these same capabilities generate significant risk. Health data is universally recognised as among the most sensitive categories of personal information, and its aggregation, linkage, and algorithmic analysis at population scale creates possibilities for surveillance, discrimination, and control that were simply not feasible under earlier, manual systems. Digital contact-tracing applications deployed during COVID-19 drew sustained criticism from human-rights bodies

[1] Crystal Ying Chan, Bridging Primary Care Gaps: Lay and Community Health Workers in Non-Communicable Disease Prevention across the Asia-Pacific, in The Handbook of Public Health in the Asia-Pacific 1 (2026), https://link.springer.com/rwe/10.1007/978-981-97-1788-0_1-1.

[2] International Congress on Academic Medicine: 2025 Medical Education Abstracts, 16 Can Med Educ J 102 (2025).

[3] Pranav Anjaria et al., Artificial Intelligence in Public Health: Revolutionizing Epidemiological Surveillance for Pandemic Preparedness and Equitable Vaccine Access, 11 Vaccines (Basel) 1154 (2023).

and civil-society organisations for insufficient legal safeguards, disproportionate data retention, and the risk that emergency surveillance infrastructure would outlive the emergency that justified it.[4] Algorithmic bias embedded in training data has been shown to reproduce and even amplify racial, socioeconomic, and geographic health disparities, with the well-documented case of a widely used US healthcare risk-prediction algorithm that systematically underestimated the health needs of Black patients standing as a cautionary example.[5]

This research paper undertakes a structured examination of this dual character. Section 2 surveys the principal applications of AI in public health surveillance. Section 3 analyses the privacy risks these applications generate, with particular attention to the COVID-19 contact-tracing experience. Section 4 examines the fundamental rights implicated, including privacy, non-discrimination, freedom of movement, and due process. Section 5 reviews the regulatory and governance architecture that has emerged to address these risks, focusing on WHO ethics guidance and the European Union's GDPR and AI Act. Section 6 considers technical and institutional mechanisms-privacy-enhancing technologies, proportionality testing, bias audits, and participatory governance-capable of reconciling innovation with rights protection. Section 7 offers case studies, and Section 8 concludes with recommendations for policymakers, technologists, and public health authorities.

The central argument advanced here is that the relationship between AI-enabled surveillance and fundamental rights is not zero-sum. Poorly designed surveillance systems undermine both privacy and public health effectiveness, since a population that distrusts a surveillance apparatus will resist engaging with it, degrading the very data quality upon which its predictive power depends. Conversely, well-governed systems-built around data minimisation, purpose limitation, transparency, and meaningful oversight-can achieve public health objectives while sustaining the public trust that is a precondition for their long-term functioning.

This argument matters beyond the specific technologies discussed here because public health surveillance is, definitionally, a state function exercised over an entire population rather than a discrete group of consenting individuals. Unlike commercial data processing, which at least formally rests on some notion of individual consent or contractual relationship, public health

[4] Human Rights Watch, World Report 2020 | Human Rights Watch (2019), https://www.hrw.org/world-report/2020.
[5] Jeena Joseph, Algorithmic Bias in Public Health AI: A Silent Threat to Equity in Low-Resource Settings, 13 Front Public Health 1643180 (2025).

surveillance frequently proceeds on the basis of a collective justification-the protection of population health-that individuals cannot readily opt out of. This structural feature makes the governance of AI in public health surveillance a distinctive and higher-stakes problem than the governance of AI in most other sectors: the same features that make AI valuable for detecting and responding to threats at population scale-its capacity to process vast, heterogeneous data streams and to identify patterns invisible to human analysts-are precisely the features that make it capable of large-scale rights infringement if deployed without adequate safeguards. The research paper's analysis therefore proceeds from the premise that questions of technical design, legal regulation, and institutional governance cannot be treated as separable from one another, but must be addressed as a single, integrated problem of responsible public health innovation.

2. AI IN PUBLIC HEALTH SURVEILLANCE: APPLICATIONS AND INNOVATION

2.1 Outbreak Detection and Early-Warning Systems

The most publicly visible application of AI in public health surveillance is automated epidemic intelligence: systems that continuously scan structured and unstructured data sources-news media, official reports, airline itineraries, social media, and clinical data-to detect signals of emerging disease activity before formal reporting channels register them. BlueDot, a Toronto-based platform founded by physician-researcher Kamran Khan, combines natural language processing with human epidemiological review to monitor more than one hundred and ninety diseases worldwide; it is credited with alerting clients to an anomalous pneumonia cluster in Wuhan on 30 December 2019, several days before the World Health Organization's public statement.[6] HealthMap, developed at Boston Children's Hospital, performs a similar function using a fully automated pipeline that aggregates online news, official health-agency reports, and social media to visualise outbreaks by location, time, and pathogen.[7] Other systems-ProMED-mail, the WHO's Epidemic Intelligence from Open Sources (EIOS), Metabiota, and EPIWATCH-occupy varying points on a spectrum from fully automated to human-moderated.[8]

A systematic review of AI-based early-warning systems found that such platforms reduce

[6] BlueDot: The World's Most Trusted Infectious Disease Intelligence, BlueDot, https://bluedot.global/
[7] Pranav Anjaria et al., Artificial Intelligence in Public Health: Revolutionizing Epidemiological Surveillance for Pandemic Preparedness and Equitable Vaccine Access, 11 Vaccines (Basel) 1154 (2023).
[8] International Congress on Academic Medicine, supra note 2.

detection latency substantially relative to traditional passive surveillance, though the review also cautioned that many of these systems remain proprietary, are not independently validated, and depend on human epidemiologists to contextualise and verify machine-generated alerts. This human-in-the-loop design is itself an important governance feature: it reduces the risk that false positives generated by noisy internet data trigger disproportionate public health responses, while preserving the speed advantage that automated scanning provides.

2.2 Predictive Modelling and Disease Forecasting

Beyond detection, machine-learning models are increasingly used to forecast the trajectory of outbreaks, incorporating epidemiological, climatic, and mobility data. Dengue and influenza forecasting have been particular areas of methodological development, with spatiotemporal models combining case data with meteorological variables to anticipate outbreak hotspots weeks in advance (Frontiers in Public Health, 2025b). During COVID-19, predictive models were used by health authorities to project hospital-bed and intensive-care demand, informing decisions about resource allocation, vaccine distribution, and the timing of non-pharmaceutical interventions.[9]

2.3 Syndromic Surveillance and Digital Epidemiology

AI systems also support syndromic surveillance-the near-real-time monitoring of clinical and behavioural indicators, such as emergency-department chief complaints, pharmacy sales, search-engine queries, and social media posts, to detect anomalous patterns that may indicate an emerging health threat before formal diagnoses are confirmed. This form of "digital epidemiology" extends surveillance beyond the clinical encounter into everyday digital behaviour, a shift that substantially expands both the analytic power and the privacy footprint of public health systems.

2.4. Resource Allocation, Vaccine Equity, and Misinformation Control

AI has also been applied to optimise vaccine distribution, identify populations at risk of under-vaccination, and detect and counter health misinformation circulating on social media platforms. These applications illustrate that the public health value of AI surveillance extends

[9] Internal Medicine [https://www utmb edu/spectre] Levine [cblevine@utmb.edu Corri, Artificial Intelligence (AI), Spectre (Nov. 25, 2024), https://www.utmb.edu/spectre/about/home/spectre-blog/2024/11/25/artificial-intelligence-(ai).

beyond outbreak detection to the broader operational challenge of translating epidemiological knowledge into equitable, timely intervention.[10]

Taken together, these applications demonstrate that AI has become embedded across the full surveillance cycle-from initial signal detection through forecasting, individual-level exposure tracking, and resource deployment. Each stage of this cycle, however, generates data-protection and rights challenges of a different character, to which the paper now turns.

3. PRIVACY RISKS AND CHALLENGES

3.1 The Special Sensitivity of Health Data

Health data occupies a privileged category in nearly every major data-protection framework. Under the European Union's General Data Protection Regulation (GDPR), health data is classified as "special category" data under Article 9, attracting the highest level of protection available in EU law and generally prohibited from processing absent an explicit legal exception, such as necessity for reasons of substantial public interest in public health.[11] This heightened sensitivity reflects the unique capacity of health information to reveal intimate facts about a person's body, genetics, sexual behaviour, mental health, and substance use-facts whose disclosure can result in stigma, discrimination in employment or insurance, and loss of social standing. When such data is processed by AI systems at population scale, three risks intensify: the sheer volume and granularity of data collected, the potential for re-identification through data linkage, and the durability of digital records relative to the transient nature of a public health emergency.

3.2 Re-identification and Aggregation Risk

A recurring finding in the literature on health data privacy is that anonymisation and aggregation, while necessary, are frequently insufficient safeguards against re-identification. Because AI systems are often designed to draw connections across disparate data sources-combining location data, social media activity, clinical records, and demographic information-the aggregation of ostensibly anonymised data streams can enable the reconstruction of individual identities, particularly for individuals in sparsely populated areas or with unusual

[10] SSG, Generative AI in Public Health: Shaping the Future of Healthcare Innovation, SSG, LLC (Feb. 20, 2025), https://www.ssg-llc.com/generative-ai-in-public-health-shaping-the-future-of-healthcare-innovation/.
[11] Your AI Assistant for Every Step of Care | Tandem Health, (Mar. 10, 2026), https://tandemhealth.ai/.

movement or health patterns. Contact-tracing applications, syndromic surveillance systems drawing on search and social media data, and epidemic-intelligence platforms scanning open-source information all present this risk to varying degrees.

3.3 Contact Tracing as a Case Study in Architectural Risk

The COVID-19 contact-tracing experience offers the clearest illustration of how technical design choices translate directly into privacy outcomes. A systematic content analysis of contact-tracing applications found substantial variation in the permissions such apps requested—including geolocation, Bluetooth, camera, and contacts access—and in the clarity, completeness, and readability of their privacy policies, with many applications requesting more data than was strictly necessary for their stated public health purpose.[12] Centralised architectures, in which proximity or location data is transmitted to and stored on government or third-party servers, create a single point of aggregation vulnerable to both authorised mission creep and unauthorised breach; decentralised architectures reduce this risk by keeping matching computations on-device, but at some cost to public health authorities' ability to conduct rich epidemiological analysis of transmission networks.[13] The United Kingdom's experience is instructive: the initial NHS contact-tracing application pursued a centralised model that drew sustained criticism from parliamentary committees, the Information Commissioner's Office, and independent legal scholars over its compliance with proportionality and necessity requirements under the European Convention on Human Rights, ultimately prompting a shift toward the decentralised Apple-Google framework.[14]

3.4 Function Creep and Mission Creep

A further, more insidious risk is "function creep"-the gradual repurposing of data or infrastructure collected for one purpose (pandemic response) toward another (law enforcement, immigration control, or commercial surveillance) without renewed public consent or legal authorisation. Human Rights Watch documented concerns that surveillance infrastructure built for contact tracing could be redirected toward policing and immigration enforcement,

[12] Marco Bardus et al., Data Management and Privacy Policy of COVID-19 Contact-Tracing Apps: Systematic Review and Content Analysis, 10 JMIR Mhealth Uhealth e35195 (2022).
[13] Katarzyna Kolasa et al., State of the Art in Adoption of Contact Tracing Apps and Recommendations Regarding Privacy Protection and Public Health: Systematic Review, 9 JMIR Mhealth Uhealth e23250 (2021).
[14] Taylor & Francis by Informa, Taylor & Francis - Fostering Human Progress through Knowledge, https://taylorandfrancis.com/

particularly in jurisdictions where police and security services operate with limited independent oversight.[15] The organisation also noted that privacy-preserving technical protocols, however well designed, cannot fully mitigate this risk where the surrounding legal and institutional environment lacks robust safeguards against repurposing, and that low-cost smartphones disproportionately used by lower-income populations often carry unpatched security vulnerabilities that undermine the practical privacy guarantees such protocols promise.

3.5 Private-Sector Involvement and Public-Private Data Flows

Much AI-enabled surveillance infrastructure is developed, hosted, or operated by private technology companies under contract with public health authorities. The International Bar Association's analysis of contact-tracing apps from a business and human-rights perspective observed that the public-private character of these partnerships distributes responsibility for rights protection across both government and corporate actors, and argued that participating businesses bear an independent obligation-exceeding mere compliance with national law-to evaluate and mitigate adverse human-rights impacts, including through leverage exercised over government partners (International Bar Association, 2020). This dynamic complicates accountability: individuals harmed by a surveillance system may struggle to determine whether responsibility lies with the commissioning public authority, the private technology vendor, or both.

3.6 Security Vulnerabilities and Data Breach Risk

Finally, the centralisation of large, linked health datasets for AI training and analysis creates an attractive target for malicious actors. Because health datasets are commercially valuable and personally sensitive, breaches carry disproportionate harm relative to breaches of less sensitive data categories. The technical literature on privacy-enhancing technologies-reviewed in Section 6-has emerged substantially in response to this risk, seeking to enable the collaborative, large-scale data analysis that AI models require without centralising raw personal data in a single vulnerable repository.[16]

[15] Human Rights Watch | Defending Human Rights Worldwide, https://www.hrw.org/
[16] Xin Gu et al., A Review of Privacy Enhancement Methods for Federated Learning in Healthcare Systems, 20 Int J Environ Res Public Health 6539 (2023).

4. FUNDAMENTAL RIGHTS AT STAKE

4.1 The Right to Privacy

The right to privacy is protected under international human-rights instruments including Article 17 of the International Covenant on Civil and Political Rights and Article 8 of the European Convention on Human Rights (ECHR), the latter of which permits interference with private life only where such interference is in accordance with law, pursues a legitimate aim, and is "necessary in a democratic society"-a standard generally understood to require that any interference be proportionate to the aim pursued. Legal analyses of COVID-19 contact-tracing applications repeatedly invoked this framework, emphasising that emergency public health measures, even where justified under derogation provisions such as Article 15 ECHR, must remain strictly limited to what the exigencies of the situation require and must not persist beyond the emergency that justifies them.[17] This "legality, necessity, and proportionality" triad has become the dominant analytical framework applied by courts, data-protection authorities, and human-rights bodies evaluating AI-enabled surveillance measures.

4.2 Non-Discrimination and the Right to Equal Treatment

AI surveillance systems risk violating the right to non-discrimination where their design, training data, or deployment produces systematically worse outcomes for particular populations. A widely cited example is a healthcare risk-prediction algorithm used across US hospitals that relied on healthcare costs as a proxy for healthcare need, a choice that-because Black patients historically incurred lower healthcare spending for reasons rooted in unequal access rather than lower need-caused the algorithm to systematically underestimate the health needs of Black patients relative to white patients with comparable underlying illness.[18] Analogous risks arise in public health AI more broadly: training datasets drawn disproportionately from urban, well-resourced healthcare systems can produce models that perform poorly when deployed in rural, low-income, or historically marginalised communities, systematically underdiagnosing or misclassifying members of underrepresented groups. A formal comment in the public health ethics literature has argued that regulatory efforts to prevent such discriminatory outcomes should not be diluted in the name of accelerating

[17] Global Insight - 2020 Archive, https://www.ibanet.org/Publications/global-insight-2020
[18] Marshall H. Chin et al., Guiding Principles to Address the Impact of Algorithm Bias on Racial and Ethnic Disparities in Health and Health Care, 6 JAMA Netw Open e2345050 (2023).

innovation, since compromising equity to expedite deployment concentrates benefits among already-advantaged populations while displacing harm onto the vulnerable.

4.3 Freedom of Movement and Assembly

Location-based surveillance technologies, including contact-tracing applications and mobility-analytics systems used to monitor compliance with quarantine or lockdown measures, implicate the right to freedom of movement. Where such systems create records of an individual's associations and movements that could later be used to identify participants in protests, religious gatherings, or other constitutionally protected activities, they raise concerns for freedom of assembly and association as well, particularly in jurisdictions with weak rule-of-law protections.

5. REGULATORY AND GOVERNANCE FRAMEWORKS

5.1 The World Health Organization's Ethics and Governance Guidance

In June 2021, the World Health Organization published Ethics and Governance of Artificial Intelligence for Health, the first global guidance document of its kind, developed over an eighteen-month process by a multidisciplinary expert group spanning ethics, law, digital technology, and human rights. The guidance articulates six consensus principles intended to anchor AI governance in the health sector: protecting human autonomy; promoting human well-being, safety, and the public interest; ensuring transparency, explainability, and intelligibility; fostering responsibility and accountability; ensuring inclusiveness and equity; and promoting AI that is responsive and sustainable. The guidance explicitly frames these principles within a human-rights paradigm, stating that ethical considerations and human rights must be placed at the centre of the design, development, and deployment of AI technologies for health, and that existing biases encoded in healthcare data-rooted in race, ethnicity, age, and gender-must be actively addressed rather than passively inherited by new AI systems. In 2025, WHO extended this framework with supplementary guidance addressing large multi-modal models, reflecting the rapid emergence of generative AI tools in clinical and public health settings and the distinct governance challenges such tools pose, including the risk of AI-generated misinformation and clinical hallucination.[19]

[19] Kenneth W. Goodman et al., Global Health and Big Data: The WHO's Artificial Intelligence Guidance, 119 S

While influential as a normative reference point, WHO guidance is non-binding "soft law." Its effectiveness depends on translation into enforceable domestic law and institutional practice-a translation that has proceeded unevenly across jurisdictions.

5.2 The European Union's General Data Protection Regulation

The GDPR remains the most consequential binding legal framework governing health data processing by AI systems, both within the EU and, through its extraterritorial reach, for organisations elsewhere that process the data of EU residents. Article 9 designates health data as a special category subject to a general prohibition on processing, lifted only where a specific exception applies-most relevantly for public health surveillance, the exception for processing necessary for reasons of substantial public interest in the area of public health.[20] Article 35 requires data controllers to conduct Data Protection Impact Assessments (DPIAs) for processing likely to result in high risk to individuals' rights and freedoms, a requirement squarely applicable to large-scale AI-driven health surveillance systems.[21] Article 22 provides a qualified right against solely automated decision-making with legal or similarly significant effects, while Articles 15 through 21 establish the broader architecture of data-subject rights-access, rectification, erasure, restriction, portability, and objection-that individuals may invoke against AI systems processing their health data.

5.3 The EU Artificial Intelligence Act

Regulation (EU) 2024/1689, the AI Act, entered into force on 1 August 2024 as the first comprehensive, horizontal legal framework for AI regulation globally, establishing a risk-based classification system spanning prohibited practices, high-risk systems, limited-risk systems subject to transparency obligations, and minimal-risk systems.[22] AI systems that pose significant risks to health, safety, or fundamental rights-a category that encompasses most AI applications in public health surveillance and clinical decision support-are classified as high-risk and subject to a demanding compliance regime, including risk-management systems,

Afr J Sci 14725 (2023).

[20] Tandem Health Alternative: Comparison and Review 2026, Heidi AI, https://www.heidihealth.com/blog/tandem-health-alternative

[21] ScienceDirect.Com | Science, Health and Medical Journals, Full Text Articles and Books., https://www.sciencedirect.com/

[22] Commission Work Programme 2025 - European Commission, (Feb. 11, 2025), https://commission.europa.eu/strategy-and-policy/strategy-documents/commission-work-programme/commission-work-programme-2025_en.

requirements for high-quality and representative training data, technical documentation, human-oversight mechanisms, and post-market monitoring. For certain high-risk systems, the Act additionally requires deployers to conduct a Fundamental Rights Impact Assessment (FRIA) prior to deployment, a requirement layered atop, but analytically distinct from, the GDPR's DPIA obligation. Notably, Article 10(5) of the AI Act creates a targeted exception permitting the processing of special-category data under GDPR Article 9 where strictly necessary for bias monitoring, detection, and correction in high-risk AI systems-reflecting a deliberate legislative judgment that some limited processing of sensitive attributes is necessary to prevent the very discriminatory outcomes discussed in Section 4.

5.4 The European Health Data Space and the Broader EU Architecture

The European Health Data Space (EHDS) Regulation, which entered into force in 2025, complements the GDPR and AI Act by establishing a governed infrastructure for the secondary use of health data-that is, use of health data for research, innovation, and AI model training beyond the original clinical purpose for which it was collected-requiring that such secondary use occur through supervised public or quasi-public infrastructure rather than unregulated bilateral data transfers.[23] Collectively, the GDPR, AI Act, and EHDS constitute a layered regulatory architecture in which data protection, product safety, and fundamental-rights protection operate as parallel, mutually reinforcing-though not always perfectly harmonized-regimes.

6. TOWARD RESPONSIBLE AI SURVEILLANCE: TECHNICAL AND INSTITUTIONAL MECHANISMS

6.1 Privacy-Enhancing Technologies

A significant body of technical research has developed methods that allow AI models to be trained on sensitive health data without centralising that data in a single, vulnerable repository. Federated learning enables multiple institutions-hospitals, public health agencies, or research centres-to train a shared model collaboratively while retaining raw data locally, with only model updates, rather than underlying records, transmitted to a central server.[24] Differential

[23] 2026 Atlas: Freedom and Prosperity around the World, Atlantic Council, https://www.atlanticcouncil.org/programs/freedom-and-prosperity-center/2026-atlas-freedom-and-prosperity-around-the-world/

[24] Considerations for Patient Privacy of Large Language Models in Health Care: Scoping Review, 27 Journal of

privacy complements this approach by introducing calibrated statistical noise into data or model outputs, providing a mathematically quantifiable guarantee that the presence or absence of any single individual's data cannot be reliably inferred from the model's output. Recent technical work has explored combining federated learning with differential privacy and homomorphic encryption for applications including cardiovascular risk modelling and biomedical image classification, generally finding that these techniques can preserve strong privacy guarantees while incurring measurable, though often manageable, costs in model accuracy and computational overhead. These technologies do not eliminate privacy risk entirely-the underlying trade-off between privacy protection and model utility persists-but they represent a meaningful technical complement to legal safeguards, particularly for cross-institutional and cross-border public health data collaboration.

6.2 Data Minimisation, Purpose Limitation, and Sunset Clauses

Beyond technical tools, institutional design choices substantially affect the rights impact of AI surveillance systems. Data minimization-collecting only the data strictly necessary for a defined public health purpose-and purpose limitation-prohibiting repurposing of data beyond that original purpose without fresh legal authorization-were identified by the UK Information Commissioner's Office as core principles for COVID-19 contact-tracing applications, alongside time limitation (ensuring data and infrastructure are decommissioned once the emergency has passed), secure processing, and meaningful user control. Sunset clauses, which impose automatic expiry on emergency surveillance powers and infrastructure absent explicit legislative renewal, directly address the function-creep risk discussed in Section 3.4 and were a central recommendation of the UK Joint Committee on Human Rights in its review of contact-tracing.[25]

6.3 Transparency, Explainability, and Human Oversight

WHO's ethics guidance identifies transparency and explainability as foundational principles, requiring that AI systems used in health be intelligible not only to technical experts but to the healthcare workers, patients, and communities affected by their use. The EU AI Act operationalises this principle through binding obligations: Article 14 requires that high-risk AI

Medical Internet Research (2025), https://www.sciencedirect.com/org/science/article/pii/S1438887125015055.

[25] Claudia Pagliari, The Ethics and Value of Contact Tracing Apps: International Insights and Implications for Scotland's COVID-19 Response, 10 J Glob Health 020103.

systems be designed to allow for effective human oversight capable of preventing or minimising risks to health, safety, or fundamental rights, while Article 13 requires that deployers be informed when they are subject to a high-risk AI system's determinations.[26] In the epidemic-intelligence context, the "human-in-the-loop" design adopted by systems such as BlueDot-in which automated signal detection is routed through human epidemiological review before triggering client alerts-illustrates a practical application of this principle, reducing the risk that noisy or erroneous machine-generated signals translate directly into disproportionate public health responses.[27]

6.4 Algorithmic Bias Auditing and Equity-Centred Design

Addressing the discrimination risks documented in Section 4.2 requires structured intervention across the algorithm life cycle. A conceptual framework developed by a panel convened by the US Agency for Healthcare Research and Quality identifies five sequential phases-problem formulation, data selection and management, algorithm development, algorithm validation, and deployment and monitoring-at each of which equity considerations must be deliberately incorporated, alongside cross-cutting commitments to transparency, explainability, authentic community engagement, and accountability for equitable outcomes. Technical fairness auditing tools, including systematic profiling of underrepresentation, label bias, and proxy variables in training data, provide complementary mechanisms for detecting discriminatory risk before deployment.[28]

7. CASE STUDIES

7.1 BlueDot and Human-in-the-Loop Epidemic Intelligence

BlueDot's detection of the Wuhan pneumonia cluster in December 2019 is frequently cited as a landmark demonstration of AI's public health value, and the case merits attention precisely because it illustrates both the promise and the limits of automated surveillance. The system's underlying methodology-natural language processing applied to hundreds of thousands of open-source data streams, combined with airline ticketing data to model geographic spread-

[26] IAPP Asia Forum 2026: Privacy | AI Governance | Cybersecurity Law | IAPP, IAPP.org, https://iapp.org/conference/iapp-asia-forum

[27] Bryan Tegomoh, The Public Health AI Handbook: Evaluating AI Tools for Public Health Practice (2025), https://publichealthaihandbook.com/.

[28] Marshall H. Chin et al., Guiding Principles to Address the Impact of Algorithm Bias on Racial and Ethnic Disparities in Health and Health Care, 6 JAMA Netw Open e2345050 (2023).

enabled detection roughly nine days ahead of the WHO's public statement. Yet BlueDot's own account of its process emphasises that automated detection was followed by human epidemiological review before any alert was issued to clients, reflecting a deliberate design choice to treat AI as an instrument for surfacing signals for expert judgement rather than as an autonomous decision-maker. This case also illustrates governance limitations: BlueDot's algorithm remains proprietary and has not been independently validated, and its commercial subscription model places its intelligence outside the reach of many lower-resourced public health departments, raising equity concerns about differential access to AI-enabled early warning across wealthy and low-resource health systems.[29]

7.2 Algorithmic Bias in Healthcare Risk Prediction

The widely documented case of a commercial healthcare risk-prediction algorithm used across major US health systems provides a cautionary account of how facially neutral design choices can produce systematically discriminatory outcomes. The algorithm employed historical healthcare spending as a proxy variable for underlying healthcare need in order to identify patients who would benefit from additional care-management resources; because Black patients, as a consequence of long-standing unequal access to care, historically incurred lower healthcare costs than white patients with comparable levels of underlying illness, the algorithm systematically assigned Black patients lower risk scores than their actual health needs warranted, resulting in reduced referral to beneficial care-management programmes. Subsequent research demonstrated that recalibrating the algorithm around a more direct measure of health need substantially reduced this disparity, illustrating that algorithmic bias, once identified, is often technically remediable-but only where systems are subject to sufficiently rigorous, equity-focused validation to surface such bias in the first place.[30]

8. DISCUSSION AND RECOMMENDATIONS

The evidence surveyed in this research paper supports several converging conclusions relevant to policymakers, public health authorities, and AI developers.

[29] CNBC com staff, These Are the 2020 CNBC Disruptor 50 Companies, CNBC (June 16, 2020), https://www.cnbc.com/2020/06/16/meet-the-2020-cnbc-disruptor-50-companies.html.

[30] Press Release | UN Launches AI Governance for Humanity Lab in Valencia | Office for Digital and Emerging Technologies, https://www.un.org/digital-emerging-technologies/content/press-release-un-launches-ai-governance-humanity-lab-valencia

First, the architecture of AI surveillance systems-not merely their stated purpose-determines their rights impact. Centralised versus decentralised data processing, the scope of data collected relative to what is strictly necessary, and the presence or absence of human review at key decision points are design choices with direct legal and ethical consequences, and should be treated as such from the earliest stages of system design rather than addressed retroactively through post-hoc privacy policies.

Second, sunset provisions and independent oversight are essential safeguards against function creep. Emergency public health powers, including the surveillance infrastructure built to exercise them, should be subject to automatic expiry absent explicit, reasoned legislative renewal, together with independent review mechanisms capable of assessing whether continued deployment remains necessary and proportionate.

Third, algorithmic bias must be treated as a first-order public health equity issue, not a secondary technical concern. Given the demonstrated capacity of biased algorithms to reproduce and amplify existing health disparities, equity-focused validation should be embedded across the full algorithm life cycle, from problem formulation through post-deployment monitoring, with particular attention to the representativeness of training data drawn from low-resource and historically underserved settings.

Fourth, privacy-enhancing technologies such as federated learning and differential privacy, while not a complete substitute for legal and institutional safeguards, offer a valuable technical complement capable of enabling collaborative, large-scale public health analysis while substantially reducing the risk associated with centralising raw sensitive data.

Fifth, regulatory harmonisation remains an unfinished project. The EU's layered GDPR-AI Act-EHDS architecture represents the most developed regulatory model currently in operation, but its complexity-and documented areas of overlap and divergence between GDPR and AI Act impact-assessment requirements-illustrates the difficulty of achieving coherent governance even within a single, well-resourced jurisdiction. Other jurisdictions, operating with more fragmented sectoral or emerging frameworks, face a correspondingly greater governance challenge, underscoring the continued relevance of WHO's non-binding but globally applicable ethics guidance as a common reference point.

Finally, public trust should be understood as a strategic public health asset rather than a mere

by-product of good governance. Surveillance systems designed and deployed with genuine attention to privacy, equity, and rights protection are more likely to secure the sustained public engagement upon which their effectiveness ultimately depends.

These conclusions carry practical implications for the three principal audiences engaged with AI public health surveillance. For policymakers and legislators, the priority is to enact clear, binding legal frameworks-rather than relying solely on non-binding ethical guidance-that specify data-retention limits, mandate independent oversight bodies, and establish enforceable rights of redress for individuals harmed by algorithmic error or discriminatory outcomes; the EU's GDPR-AI Act architecture, whatever its residual complexity, offers the most developed template currently available for other jurisdictions to adapt. For public health authorities and practitioners, the priority is to resist the temptation to treat AI procurement as a purely technical decision, instead building institutional capacity-including in-house epidemiological, legal, and data-governance expertise-sufficient to evaluate vendor claims, demand meaningful transparency about model design and training data, and maintain genuine human oversight rather than reflexive deference to automated outputs. For AI developers and technology vendors, the priority is to internalise equity and privacy protection as design requirements from the outset of system development, rather than as compliance obligations addressed after a system's core architecture has already been fixed, recognising that retrofitting privacy and fairness protections onto an already-deployed system is both more costly and less effective than embedding them from inception.

9. CONCLUSION

Artificial intelligence has fundamentally expanded the capacity of public health authorities to detect, forecast, and respond to threats to population health, from the early identification of novel pathogens to the granular reconstruction of transmission chains during a global pandemic. This expanded capacity, however, is inseparable from an expanded capacity for privacy intrusion, discriminatory harm, and the erosion of fundamental rights, particularly where surveillance systems are designed and deployed without adequate attention to data minimisation, algorithmic bias, transparency, and independent oversight. The regulatory and technical responses that have since emerged-WHO's ethics guidance, the EU's GDPR-AI Act-EHDS framework, privacy-enhancing technologies, and equity-centred algorithm design methodologies-demonstrate that innovation and rights protection can be reconciled, but only

through deliberate institutional design rather than as an automatic consequence of technological progress. As AI capabilities continue to advance, the central challenge for public health governance is not whether to use AI in surveillance, but how to embed privacy, equity, and fundamental-rights protection into the foundations of these systems from their inception, ensuring that the pursuit of population health does not come at the unwarranted expense of the individuals that public health is ultimately meant to serve.