

Next-Gen AI-Enhanced Epidemic Intelligence Systems: Policies for Equitable, Privacy-Preserving, Sustainable & Croked Innovations for AI Applications in Infectious Diseases Surveillance

Submitted by: Anthony Li, Axby Loh, Waikit Fung

AIM

This paper proposes for the White House Office of Pandemic Preparedness and Response Policy (OPPR) to drive an interagency and public-private program to revolutionize epidemic intelligence through equitable, privacy-preserving Artificial Intelligence(AI)-enhanced epidemic intelligence, to save lives millions at a time.

PROBLEM

2. Epidemic intelligence is a critical element of pandemic preparedness and response by enabling public health agencies to control outbreaks and save lives. AI offers transformative potential for epidemic intelligence, with applications including early warning of emerging infectious diseases, public health threat assessment, predictive modeling and resource allocation.¹ To leverage this technological advancement, the US Health and Human Services (HHS) and Communicable Diseases Control and Prevention (CDC) have launched the Data Modernization Initiative (DMI), invested in the Response Ready Data Enterprise Integration (RRDEI) platform and established the Center for Forecasting and Outbreak Analytics (CFA). Refer to **ANNEX A** for an overview of existing initiatives.

3. **Gaps in Current AI Applications.** While these initiatives mark significant progress, they scratch the surface of AI's full potential in epidemic intelligence, with many AI applications not fully probed:

- A. **Function:** Currently, efforts such as the Center for Forecasting and Outbreak Analytics (CFA) and Insight Net primarily focus on disease forecasting. However, the broader epidemic intelligence process consists of numerous activities where AI could be impactful such as prioritizing diagnostic tests at points-of-entry² and enabling community-driven epidemic intelligence (see *Para. 9* below for details).
- B. **Machine learning type:** Majority of existing epidemic intelligence systems employ supervised (e.g., CDC's FluSight) and unsupervised (e.g., the Global Public Health Intelligence Network) learning to form predictions³. The application of Reinforcement Learning and Deep Learning methods remain limited and hold immense potential for detecting complex patterns in large volumes of unstructured data.

4. **Barriers to Advancing AI-enhanced Epidemic Intelligence.** To effectively implement AI epidemic intelligence systems across the entire intelligence process, several barriers need to be addressed:

- A. **Data limitations:** The effectiveness of AI tools can be constrained by data quantity and quality. While the DMI and RRDEI have integrated diverse data sources to support outbreak response, critical data for epidemic intelligence—such as information from One Health agencies, open-source platforms, and commercial or administrative data—remain unincorporated. Data integration of data sources outside of health offers untapped opportunities for deep learning

¹ Brownstein, J. S., Rader, B., Astley, C. M., & Tian, H. (2023). Advances in artificial intelligence for infectious-disease surveillance. *New England Journal of Medicine*, 388(17), 1597-1607.

² Bastani H, Drakopoulos K, Gupta V, et al. Efficient and targeted COVID-19 border testing via reinforcement learning. *Nature* 2021;599:108-13.

³ Anjaria, P., Asediya, V., Bhavsar, P., Pathak, A., Desai, D., & Patil, V. (2023). Artificial Intelligence in Public Health: Revolutionizing Epidemiological Surveillance for Pandemic Preparedness and Equitable Vaccine Access. *Vaccines*, 11(7), 1154. <https://doi.org/10.3390/vaccines11071154>

neural networks to grok epidemic intelligence systems with higher sensitivity and greater throughput.⁴

- B. **Talent shortage:** Developing AI-enhanced epidemic intelligence systems requires professionals with specialized expertise in both AI and infectious disease. The current talent gap in these fields risks delaying federal efforts to deploy AI-enhanced epidemic intelligence capabilities.
- C. **Resource constraint:** AI development is resource-intensive, relying heavily on large datasets and substantial computing power. Competition for computing resources, such as graphical processing units (GPUs), pose significant risks to the AI development process.
- D. **Ethical concerns:** Pressing ethical issues must be addressed to maintain the public trust essential for effective deployment:
 - i. **AI Model Bias and Discrimination:** AI models can inadvertently perpetuate human biases, and lead to discrimination, particularly when trained on health datasets that lack representation of vulnerable and marginalized populations.
 - ii. **Data Privacy and Patient Confidentiality:** AI tools pose risks to patient privacy by memorizing and inadvertently disclosing sensitive health information. Additionally, the use of patient data in AI model development without explicit consent can infringe on patient confidentiality.

5. To address these challenges, our recommendations are detailed in [paragraphs 9-12](#). Prior to that, we outline key considerations in the following section, [paragraphs 6-8](#).

CONSIDERATIONS

- 6. **Ensuring Ethical Development and Deployment.**
 - A. **Uphold equity.** AI tools can inadvertently reflect societal biases if the data used for training is not representative, leading to the unfair targeting of specific demographics and compromising equity.
 - B. **Safeguard privacy and confidentiality.** The development of AI-enhanced epidemic intelligence systems requires the collection of extensive datasets, some of which may contain personal information. Policies supporting the use of AI for epidemic intelligence must therefore incorporate robust privacy safeguards and ensure compliance with the Health Insurance Portability and Accountability Act.
 - C. **Ensure safety.** AI systems can be susceptible to cybersecurity threats. Large language models, in particular, are vulnerable to prompt injections used to reveal confidential information.
- 7. **Facilitate Sustainable innovation.** Recognizing the life-saving potential of AI-enhanced epidemic intelligence tools, regulatory measures must be carefully calibrated to avoid hindering innovation and adoption. Policies should foster a supportive environment for innovation, particularly in high-potential applications of AI in epidemic intelligence. By balancing regulatory oversight with the need for flexibility, we can encourage the development and implementation of innovative solutions that enhance public health outcomes.
- 8. **Public Private Partnerships (PPP).** Due to the considerable amount of resources (human, data and compute resources) required to develop AI-enhanced epidemic intelligence systems, it will be important for the government to collaborate with industry, academia, and non profit partners. For instance, industry partners like NVIDIA and Meta can provide support with GPUs and state of the art Large Language Models (LLMs), public health academics can help to assess public health impact, and promote equity by engaging with LGBTQ+ and Black communities. This collaborative approach will enhance the effectiveness and inclusivity of AI-enhanced epidemic intelligence systems.

POLICY RECOMMENDATION

- 9. **Advance AI-Enhanced Epidemic Intelligence Systems.** We recommend that the Office of Public Health Preparedness and Response (OPPR) support epidemic intelligence agencies in adopting AI technologies while safeguarding equity and privacy. OPPR is recommended to enable the development of:
 - A. **Risk-informed testing.** The government of Greece implemented "Eva," an AI tool utilized at points-of-entry to identify travelers to be tested for COVID-19 using polymerase-chain-reaction

⁴ Liu, Z, Kitouni, O, Nolte, NS. Towards understanding grokking: an effective theory of representation learning. Adv Neural Informat Proc Syst. 2022; 35:34651-34663

(PCR). The test results were then fed back anonymously to enable Eva to continuously refine its predictive accuracy. This approach demonstrated an accuracy improvement of 25% to 45% over conventional testing methods, which often targeted individuals from countries with high case rates⁵.

- B. **Contact tracing.** Researchers at the University of Pittsburgh developed the Enhanced Detection System for Healthcare-Associated Transmission (EDS-HAT) to trace hospital-acquired infections by leveraging AI to mine patients' electronic medical records automatically. This tool has the potential to prevent up to 40% of hospital-acquired infections and could be adapted for tracing disease transmission beyond healthcare settings.⁶
- C. **Hotspot detection.** Researchers used an artificial neural network model to identify spatial clusters and socioeconomic factors associated with tuberculosis. Such insights can guide the prioritization of resources toward vulnerable communities and geographies.⁷
- D. **Community-driven epidemic intelligence.** In Tanzania, the AfyaData⁸ application enables community health workers to report disease symptoms and cases to a centralized system. AI algorithms then analyze the data for patterns, generating alerts when unusual patterns are detected. These alerts go back to both health authorities and local communities, enabling quicker, locally tailored responses to potential outbreaks. By involving communities in AI development and data collection, public health initiatives build trust, promote public health education, and create a stronger, more resilient public health system.

10. **Establish AI-supporting Infrastructure.** The Office of Public Health Preparedness and Response (OPPR) can support the adoption of AI for epidemic intelligence:

- A. **Coordinating agency.** Designating the HHS as the lead coordinating agency responsible for driving AI adoption in epidemic intelligence among the Centers for Disease Control and Prevention (CDC), U.S. Department of Agriculture (USDA), Department of Defense (DOD), and other relevant agencies.
- B. **Data and Talent Development.** Broaden the scope of the DMI to incorporate a wider range of data pertinent to epidemic intelligence, including information from One Health agencies (e.g. USDA), open-source data (e.g. anonymized data from public social media profiles and advertising analytics), commercial (e.g. pharmacy purchases) and epidemic intelligence data in other animal species (especially where zoonotic spillovers are relevant). The CDC should continue to establish national standards for data quality and interoperability in public health epidemic intelligence. Additionally, DMI's workforce development programs should be expanded across agencies beyond the CDC.
- C. **Shared computing hardware.** Since training AI models typically requires substantial computing resources (e.g., GPUs) for a one-time use, it is cost-inefficient for each agency to maintain their own computing capabilities. A lack of adequate computing power could hinder the effectiveness of AI models. The CDC could acquire and manage the necessary computing hardware, allowing other agencies to access these resources through cloud services for the training of AI tools used in epidemic intelligence.
- D. **Horizontal privacy-preserving platforms.** To integrate a broad range of data while preserving privacy, the CDC should be tasked with developing and managing privacy-preserving platforms to support AI-enhanced epidemic intelligence. A key approach would involve federated learning as the default training method for all AI-enhanced tools. Through federated learning, data remains within the secure databases of contributing organizations, and only anonymized model parameters—not raw data—are shared to train a unified foundational AI model. Blockchain technology could further enhance security by tracking and verifying encrypted model updates from multiple sources. For daily operations, each agency can retain a local AI model, maintaining full data control at the organizational level.
- E. **Research and Development.** The OPPR can serve as the custodian of a shared fund of \$50 million dedicated to procuring computing hardware and funding projects across agencies aimed at developing AI tools for epidemic intelligence. This investment would support innovation and collaboration, driving advancements in epidemic intelligence. The OPPR can also ensure that the

⁵ Bastani H, Drakopoulos K, Gupta V, et al. Efficient and targeted COVID-19 border testing via reinforcement learning. *Nature* 2021;599:108-13.

⁶ Sundermann AJ, Miller JK, Marsh JW, et al. Automated data mining of the electronic health record for investigation of healthcare-associated outbreaks. *Infect Control Hosp Epidemiol* 2019;40:314-9

⁷ Mollalo A, Mao L, Rashidi P, Glass GE. A GIS-based artificial neural network model for spatial distribution of tuberculosis across the continental United States. *Int J Environ Res Public Health* 2019;16: 157.

⁸ Karimuribo, E. D., Mutagahywa, E., Sindato, C., Mboera, L., Mwabukusi, M., Kariuki Njenga, M., Teesdale, S., Olsen, J., & Rweyemamu, M. (2017). A Smartphone App (AfyaData) for Innovative One Health Disease Surveillance from Community to National Levels in Africa: Intervention in Disease Surveillance. *JMIR public health and surveillance*, 3(4), e94. <https://doi.org/10.2196/publichealth.7373>

intellectual property of the AI-enhanced epidemic intelligence tools created by agencies is protected, should there be commercial value.

11. **Implement Oversight Measures.** The Office of Public Health Preparedness and Response (OPPR) can ensure safety, equity, privacy, and confidentiality throughout all stages of AI development and utilization by implementing the following oversight measures:
 - A. **Equitable data collection.** The data collection process should be conducted equitably, inclusively and representatively across diverse populations. Guidelines should be established to ensure that infectious diseases data collection is sensitive to cultural considerations and linguistic diversity. Furthermore, demographic data should encompass marginalized populations, racial minorities, gender non-conforming individuals, and low-income communities.
 - B. **Data privacy.** Given the highly sensitive nature of infectious disease information—especially in cases involving high risk sexual behaviors such as those associated with HIV and monkeypox—it is imperative to minimize breaches of patient privacy. AI systems should be implemented with robust confidentiality measures, including encryption, access controls, and data minimization strategies. Data manipulated by AI systems should be de-identified to the greatest extent possible. Additionally, outputs generated by LLMs should be parsed to eliminate any personal identifiers that may inadvertently be reproduced.
 - C. **Safe and fair AI models.** Systems enhanced by large language models (LLMs) must incorporate negative prompts or mechanisms to guard against prompt injection vulnerabilities. AI models should undergo rigorous third-party evaluations and red teaming both prior to deployment and after implementation to confirm their safety, effectiveness, and freedom from unintended biases.
 - D. **AI governance.** All governmental institutions and public healthcare entities should establish in-house AI governance bodies tasked with overseeing adherence to ethical principles and existing regulations. These bodies should provide mechanisms for employees to report concerns and highlight adverse events, assess reported incidents and offer guidance on corrective actions.
 - E. **Equitable access to AI products and epidemic intelligence findings.** AI products should be made accessible to state, local, tribal, and territorial health agencies, empowering communities to monitor outbreaks effectively. Furthermore, insights generated by these tools should be shared publicly to promote health-promoting behaviors. Communities with shared epidemic intelligence findings can then be involved in providing feedback data that can refine predictive accuracy.
12. **Phased regulation to balance innovation, equity, safety and privacy.** To facilitate sustainable innovation and avoid hindering adoption, regulatory measures must be carefully calibrated through the following approach:
 - A. **AI and data governance guidelines and frameworks:** Given the emerging role of AI in epidemic intelligence, regulatory measures should be carefully calibrated to avoid stifling innovation. As a preliminary approach, implementing clear guidelines and governance frameworks can provide essential guardrails to promote equitable, fair, and safe AI use in this domain.
 - B. **AI epidemic intelligence sandbox:** Before translating the guidelines into regulations, establishing an AI epidemic intelligence sandbox offers a controlled setting to test small-scale prototypes across various states. This approach allows regulators to assess real-world impacts of AI-enhanced epidemic intelligence systems, identifying system performance, potential blind spots, and associated safety and ethical considerations.
 - C. **Transitioning from guidelines to regulations and legislation:** Insights gained from the sandbox phase will inform the development of targeted legislation, equipping enforcement agencies with the authority to ensure AI systems are designed to minimize bias and inequities while upholding safety standards.

CALL TO ACTION

13. As the United States grapples with the ongoing outbreak of measles and H5N1 epizootic, alongside global incidents of Marburg and Mpox, controlling the spread of infectious diseases has become increasingly challenging. Fortunately, AI has the potential to serve as a transformative tool, allowing the government to rapidly detect and respond to emerging epidemics. To save lives, promote equity, and safeguard privacy, we urge the Office of Public Health Preparedness and Response (OPPR) to consider our recommendations outlined in **Paras. 9 – 12.**

REFERENCES

- Anjaria, P., Asediya, V., Bhavsar, P., Pathak, A., Desai, D., & Patil, V. (2023). Artificial Intelligence in Public Health: Revolutionizing Epidemiological Surveillance for Pandemic Preparedness and Equitable Vaccine Access. *Vaccines*, 11(7): 1154.
- Bastani H, Drakopoulos K, Gupta V, et al. (2021). Efficient and targeted COVID-19 border testing via reinforcement learning. *Nature*, 599: 108-13.
- Brownstein, J. S., Rader, B., Astley, C. M., & Tian, H. (2023). Advances in artificial intelligence for infectious-disease surveillance. *New England Journal of Medicine*, 388(17): 1597-1607.
- Karimuribo, E. D., Mutagahywa, E., Sindato, C., Mboera, L., Mwabukusi, M., Kariuki Njenga, M., Teesdale, S., Olsen, J., & Rweyemamu, M. (2017). A Smartphone App (AfyaData) for Innovative One Health Disease Surveillance from Community to National Levels in Africa: Intervention in Disease Surveillance. *JMIR public health and surveillance*, 3(4): e94.
- Liu, Z, Kitouni, O, Nolte, NS. (2022). Towards understanding grokking: an effective theory of representation learning. *Adv Neural Informat Proc Syst.*, 35: 34651-34663
- Mollalo A, Mao L, Rashidi P, Glass GE. (2019). A GIS-based artificial neural network model for spatial distribution of tuberculosis across the continental United States. *Int J Environ Res Public Health*, 16: 157.
- Sundermann AJ, Miller JK, Marsh JW, et al. (2019). Automated data mining of the electronic health record for investigation of healthcare-associated outbreaks. *Infect Control Hosp Epidemiol.*, 40: 314-9

EXISTING INITIATIVES USING AI FOR EPIDEMIC INTELLIGENCE

Data Modernization Initiative (DMI): Through the DMI, the CDC is establishing new standards and regulations for facilitating the modernization of data exchange between the health care and public health sectors.

[The Response Ready Data Enterprise Integration \(RRDEI\)](#) platform integrates more than 200 data sources across federal, state, and local governments and the healthcare industry to provide timely information to support evidence-based decision-making for current and emerging public health threats. Data housed and analyzed relate to Cases, Testing, Hospitalizations, Mortality, Mobility, Schools, Therapeutics, Medical supply chains, Demographics, Local policy decisions, Social determinants of health, Healthcare system capacity.

[CDC Center for Forecasting and Outbreak Analytics \(CFA\)](#) uses data, modeling, and analytics to assess public health threats to support public health response and decision-making.

[CDC Insight Net](#) brings together more than 100 academic and private entities, and state, tribal, local, and territorial (STLT) health departments in a collaborative partnership to train, develop analytical tools, advance the analysis and use of data about infectious disease spread.

[FY2025 budget for HHS](#) allocated \$225 million to the DMI, \$60 million for RRDEI and \$50 mil for CFA and Insight Net.