

BEACON

Biothreats Emergence, Analysis, and Communications Network

Larry Madoff, MD, FACP, FIDSA
Senior Advisor

EIOS GTM Salay, Senegal 12 December 2024

 Center on Emerging Infectious Diseases



Research Powering Policy and
Action

In collaboration with:

 Rafik B. Hariri Institute for Computing
and Computational Science & Engineering





We live in an
age of
epidemics



Detection



Reporting

Challenges:

- Political transparency
- Only focused on human health
- Disparate sources
- Data sharing and interoperability
- Language barriers
- Lack of geographic scope
- Lack context and prioritization/prognostication
- Behind paywall



Response

Signal Detection

- Online news aggregators
- Eyewitness reports
- Online expert-curated discussions
- Validated official reports
- 12 languages, 24/7



HealthMap
(web scrapping)



Public Health
Partners



Individual
Communications



Pandemic LLM

- Signal Prioritization
- Translation
- Copy editing and drafting
- Inclusion of reference data
- Trained on historic data

Signal Verification & Context



- Globally based SMEs
- Network of trust
- Multidisciplinary expertise
- Many ex-ProMED
- Authoring online reports

Open Source Reports (& LLM Models)

- Near real-time
- Context + data
- All UN languages
- Practitioner discussion
- Visual & data layers

Predictive Intelligence



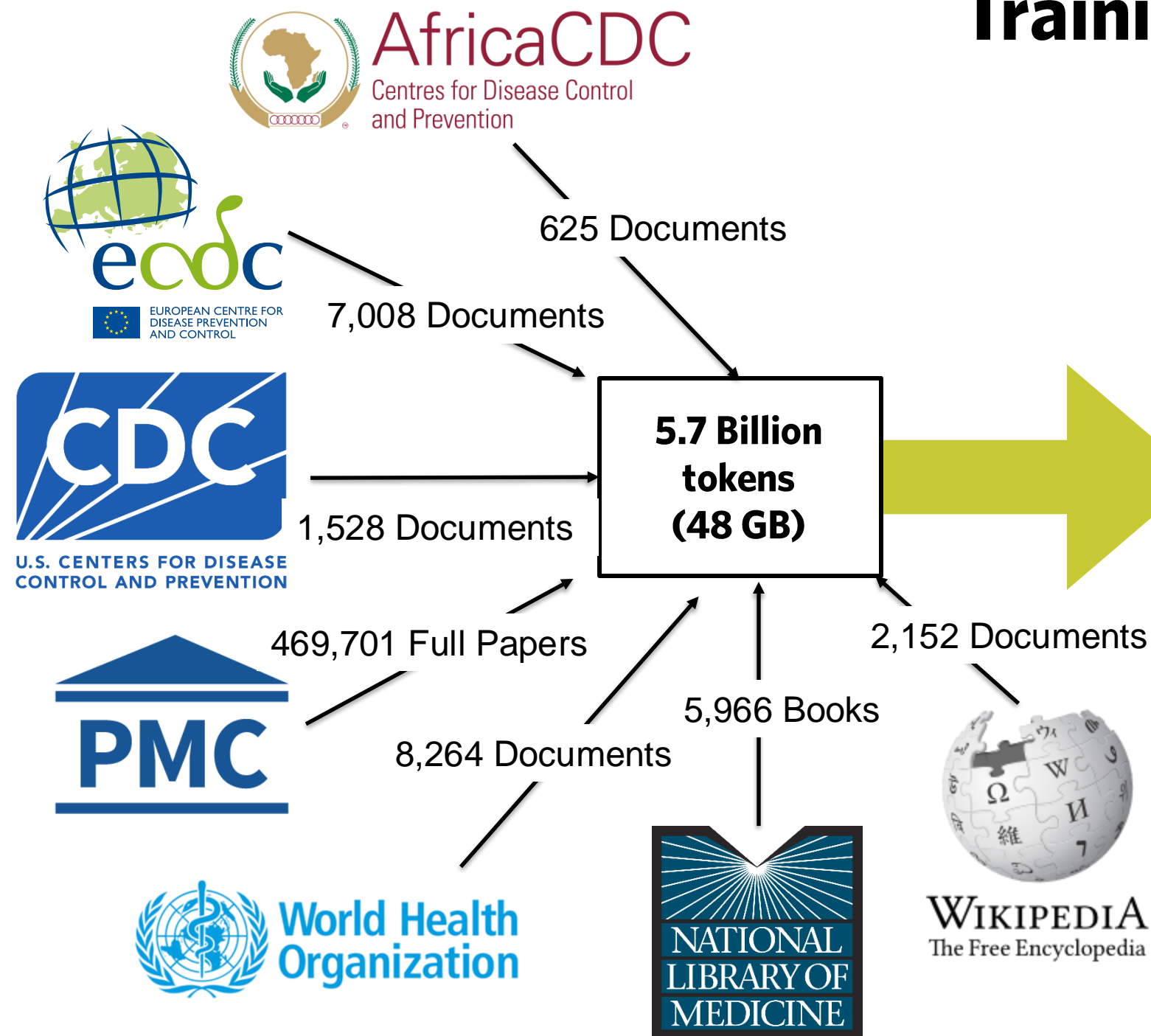
- Predicting scope, spread & severity
- Linked to SME verification
- Generated for different use cases

6-9 months

1 year

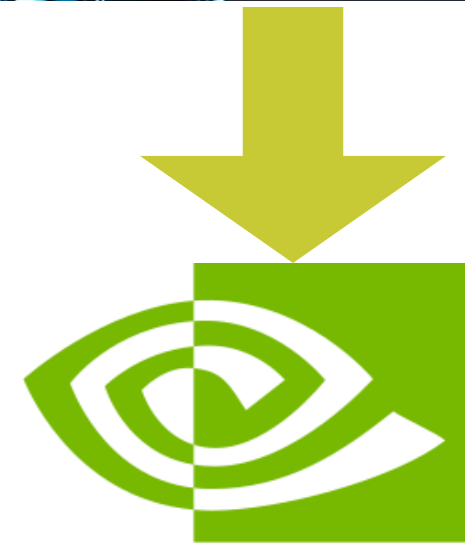
Timeline

Training Pandemic LLM



Using Publicly Available Documents

Starting from a Generic, Open-Source, Pretrained LLM



NVIDIA®

Full Parameter Fine-Tuning

- 4 A100 nodes
- 8 GPUs per node -> 32 GPUs
- 80 GB memory per GPU



Support from NSF

Support from DOE



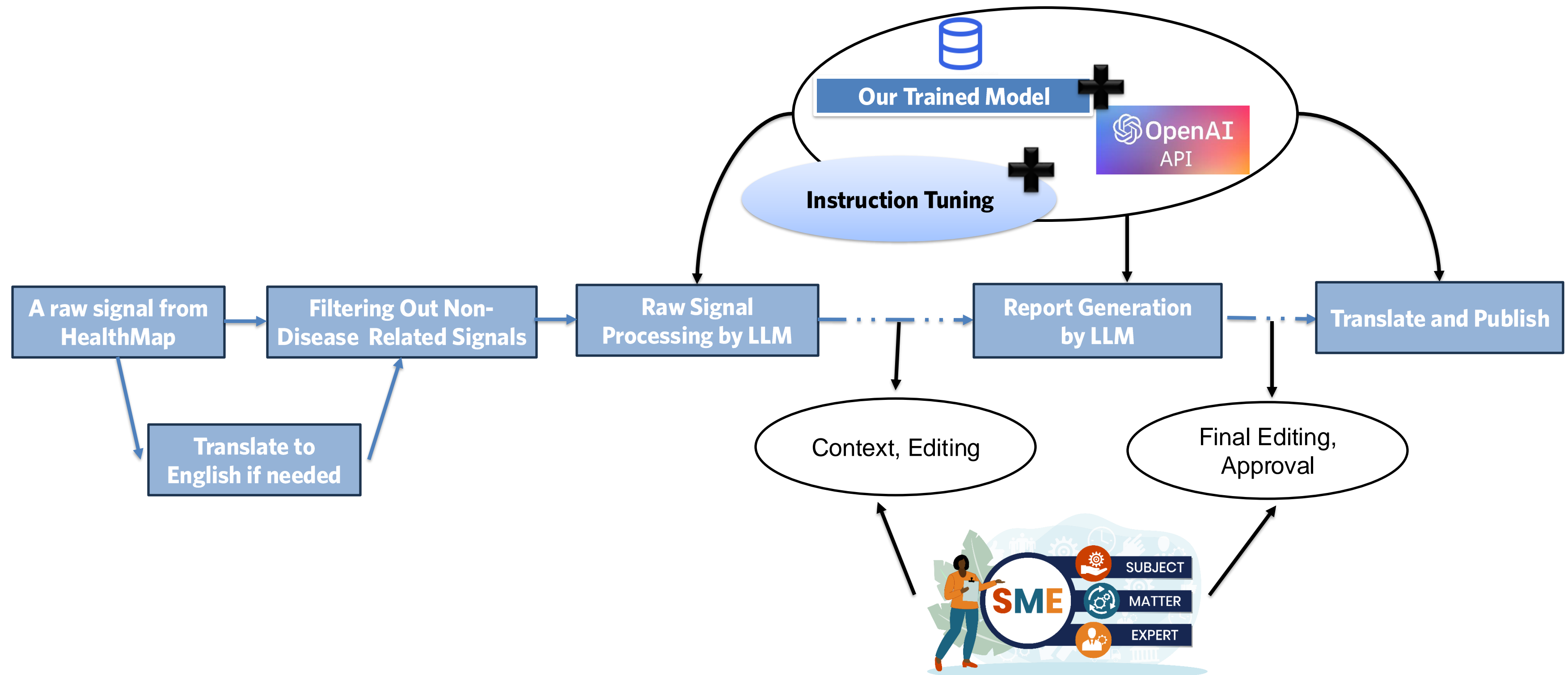
National Energy Research Scientific Computing Center

Parameter-Efficient Fine-Tuning

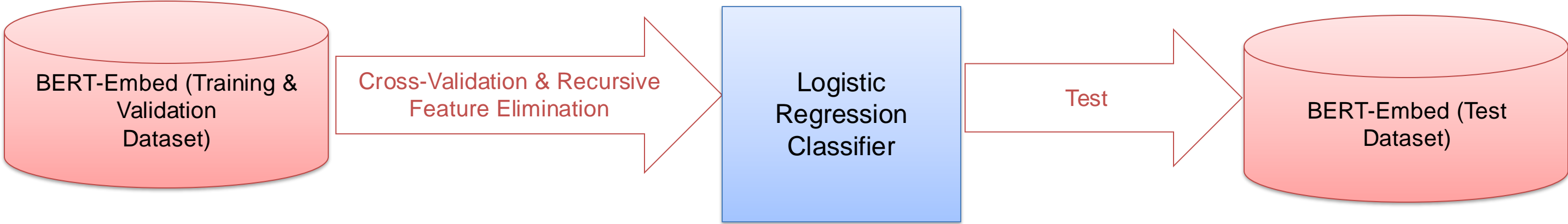
Low-Rank Adaptation

- 2 nodes
- 4 GPUs per node -> 8 GPUs
- 80 GB memory per GPU
- LLaMA 3.2 (1B parameters)
 - Total training duration: 80 hrs
- LLaMA 3.2 (3B parameters)
 - Total training duration: 240 hrs

Beacon Outbreak Reporting and Signal Prioritization Pipeline



Filtering Out Non-Disease Related Signals



- Trained on 5000 scraped HealthMap historical data on specific outbreak dates and labeled by GPT-4o.
- Reached a high AUC of 98% on test set.
- Excludes the unrelated incoming signals to reduce the load on the server and LLM processing required.

Signal Info Extraction

Basic Information	Location Details	Health Data	Risk Factors	Response Measures
<ul style="list-style-type: none">• Summary• Publish Date:• Event Date• Link	<ul style="list-style-type: none">• City• Country• Continent• Location Type	<ul style="list-style-type: none">• Diseases• Pathogens• Species• Cases• Fatalities• Symptoms• Incubation• Transmission	<ul style="list-style-type: none">• Environmental Risks• Political Risks• Spread potential	<ul style="list-style-type: none">• Prevention• Immunology• Treatment
Demographics				
<ul style="list-style-type: none">• Demographics				

Signal Processing

Recency Evaluation	Source Reliability Evaluation	Signal Quality Evaluation	Urgency Evaluation	Severity Evaluation
<ul style="list-style-type: none">• Critical• Very Recent• Recent• Moderately Recent• Not Recent• Unknown Date	<ul style="list-style-type: none">• Highly Reliable• Very Reliable• Moderately Reliable• Reliable	High Medium Low	<ul style="list-style-type: none">• Critical• High• Moderate• Low	<ul style="list-style-type: none">• Pathogen Characteristics• Clinical Impact & Disease Manifestation• Transmission Dynamics• Geographic & Demographic Factors• Healthcare System Capacity• External Risk Factors

BEACON

Biothreats Emergence, Analysis, and Communications Network

PROGRAM COMPONENTS



Web platform &
predictive
intelligence



International
Governance Board



Global
Training
Program



Community of
Practitioners

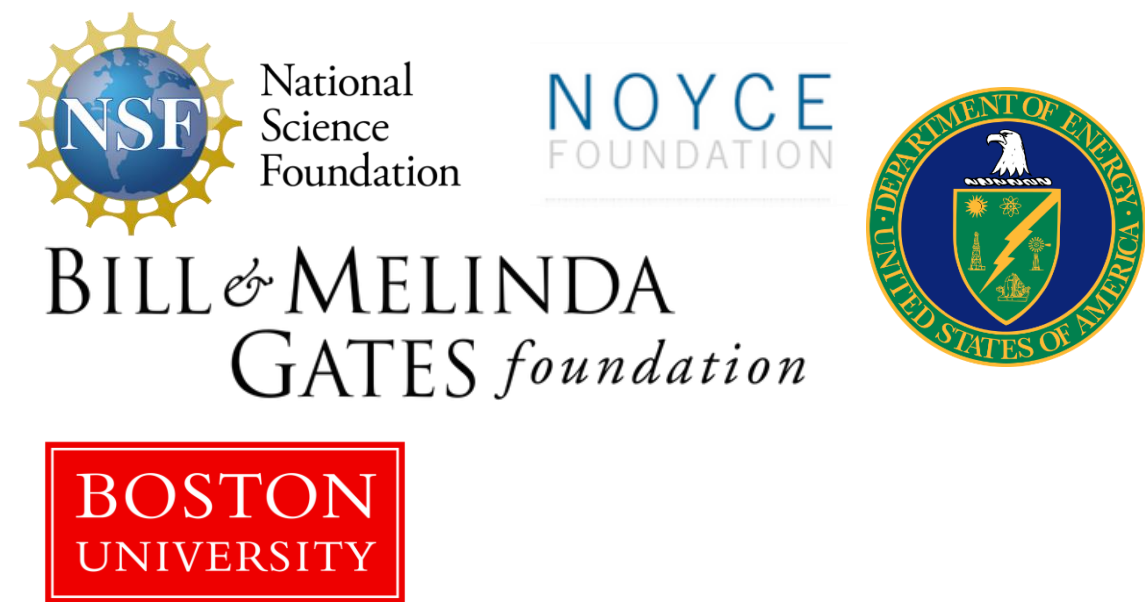
How can BEACON make a difference?

- Transparent, independent and One Health focused
- Near real time, open source, data in all UN languages
- Context regarding why a new “signal” matters to help inform non practitioners and policy makers
- Predictive intelligence to guide response scale and scope
- Serve as platform for collaboration in outbreak response
- Linkage to data, reference materials, clinical guidelines and situation reports to raise interdisciplinary situational awareness
- Ability to train global scholars on public health AI use case
- Provide data for research in public health, AI, policy and natural sciences

Who can BEACON help?



FUNDERS



TECHNICAL PARTNERS & COLLABORATORS



MORE INFORMATION



LEADERSHIP



Nahid Bhadelia, MD, MALD
Director



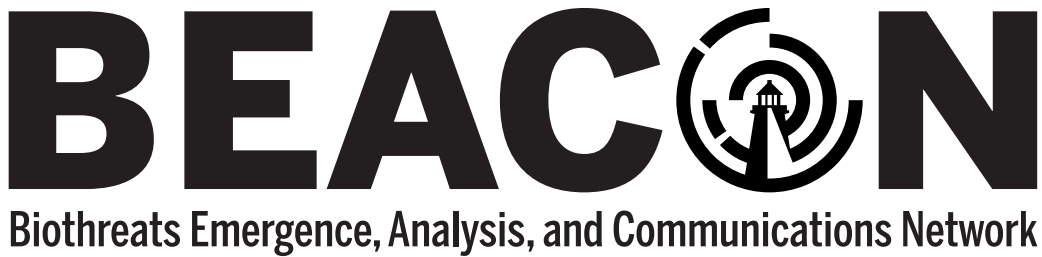
Britta Lassman, MD
Co-Director, Editor, Human moderation



John Brownstein, PhD
Co-Director, Digital Surveillance and Platform Development



Ioannis Paschalidis, PhD
Co-Director, Predictive Intelligence and Large Language Models



Buildup:

- 6-9 months for the open, AI-enabled, outbreak reporting system
- 1 additional year for predictive models and system enhancements & improvements

Steady-state:

- Continuous system and model improvements
- Building educational programs for physicians, policymakers, and stakeholders
- Advising government agencies
- Public engagement events
- Launching programs to support policy development, and the development of vaccines and therapeutics