

Pour regarder une version archivée de cette présentation, cliquer sur le lien ci-après:



<http://pho.adobeconnect.com/pquInvzhr0kf/>

Faire défiler le curseur vers le bas pour consulter les diapositives de la séance. Cet enregistrement n'est disponible qu'en anglais.

*Avertissement*

**Ce document a été créé par l'auteur et/ou par une organisation externe. Il est publié sur le site web de Santé publique Ontario (SPO) pour une utilisation publique, comme indiqué dans les conditions d'utilisation de notre site web. SPO ne possède pas le contenu de ce document. Toute application ou utilisation de l'information contenue dans ce document relève de la responsabilité de l'utilisateur. SPO décline toute responsabilité résultant d'une telle application ou utilisation.**

# Conseils utiles pour visionner la présentation :

- Le format de présentation par défaut inclut l'index de l'événement. Pour fermer l'index, cliquer sur l'icône suivant  et fermer.
- Pour visualiser la présentation en mode plein écran, cliquer sur  au haut de l'écran partagé, à droite

# Intelligence artificielle (IA) et apprentissage machine (AM) en santé publique

Laura Rosella

Stacey Fisher

Melodie Song

29 octobre 2020



**Laura C. Rosella, Ph.D.**

Professeure agrégée et titulaire de la Chaire de recherche du Canada en analyse de la santé des populations,  
École de santé publique Dalla Lana, Université de Toronto  
Directrice d'emplacement, ICES, Université de Toronto  
Professeure affiliée, Institut Vecteur

**Stacey Fisher, Ph.D.**

Titulaire d'une bourse d'apprentissage en matière d'impact sur le système de santé des IRSC, spécialisation en intelligence artificielle équitable  
Boursière de recherches postdoctorales, Santé publique Ontario  
Boursière de recherches postdoctorales, École de santé publique Dalla Lana, Université de Toronto

**Melodie Song, Ph.D.**

Titulaire d'une bourse d'apprentissage en matière d'impact sur le système de santé des IRSC, spécialisation en intelligence artificielle équitable  
Boursière de recherches postdoctorales, Santé publique Ontario  
Boursière de recherches postdoctorales VICTOIRE, École de santé publique Dalla Lana, Université de Toronto



Soutenir les applications d'**intelligence artificielle** en santé publique afin d'améliorer **l'équité en matière de santé** et de **prévenir les maladies chroniques** au sein de la population



Orienter l'élaboration  
d'une stratégie  
d'intelligence  
artificielle pour SPO



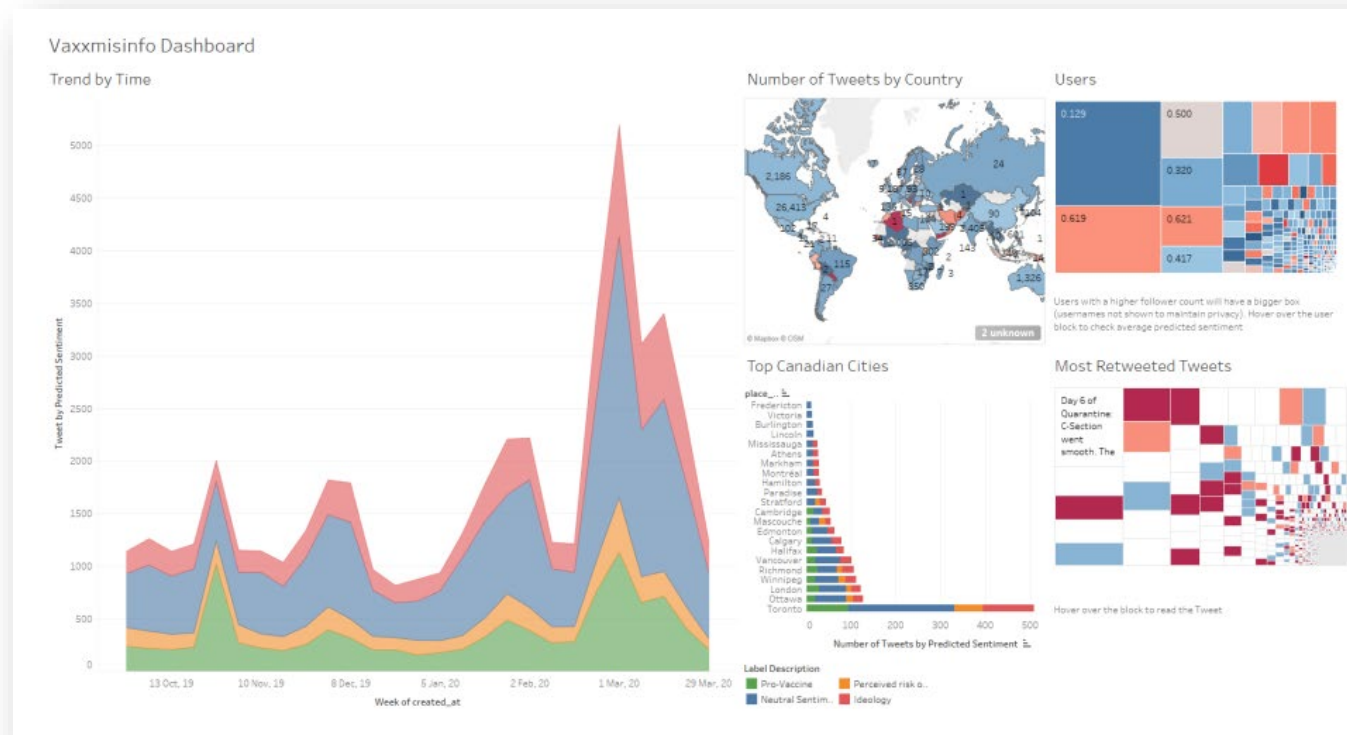
Aider les bureaux de  
santé publique à  
accroître leur capacité  
en matière d'IA/AM



Créer deux nouveaux  
modèles de prévision  
des risques à l'aide de  
méthodes d'IA/AM

# Programme de Bourses d'apprentissage en matière d'impact sur le système de santé (BAISS) – volet IA et santé publique (2019-2020)

- Examen de la portée de l'utilisation de l'IA à des fins d'immunisation
- Comités d'experts délibérants
- Tableau de bord d'apprentissage machine à intervention humaine visant à détecter la désinformation sur Twitter au sujet des vaccins



## DIVULGATIONS

- Aucun des présentateurs nommés dans ce document n'a reçu d'aide financière ni de soutien en nature de la part d'un commanditaire commercial.
- Les présentateurs n'ont aucun conflit d'intérêts potentiel à déclarer.

# Question de sondage

Sondage – Que savez-vous de l'intelligence artificielle et de l'apprentissage machine?

- A) Pratiquement rien
- B) J'ai quelques connaissances, mais je ne m'en suis jamais servi(e).
- C) J'ai un peu d'expérience.
- D) J'ai beaucoup d'expérience.



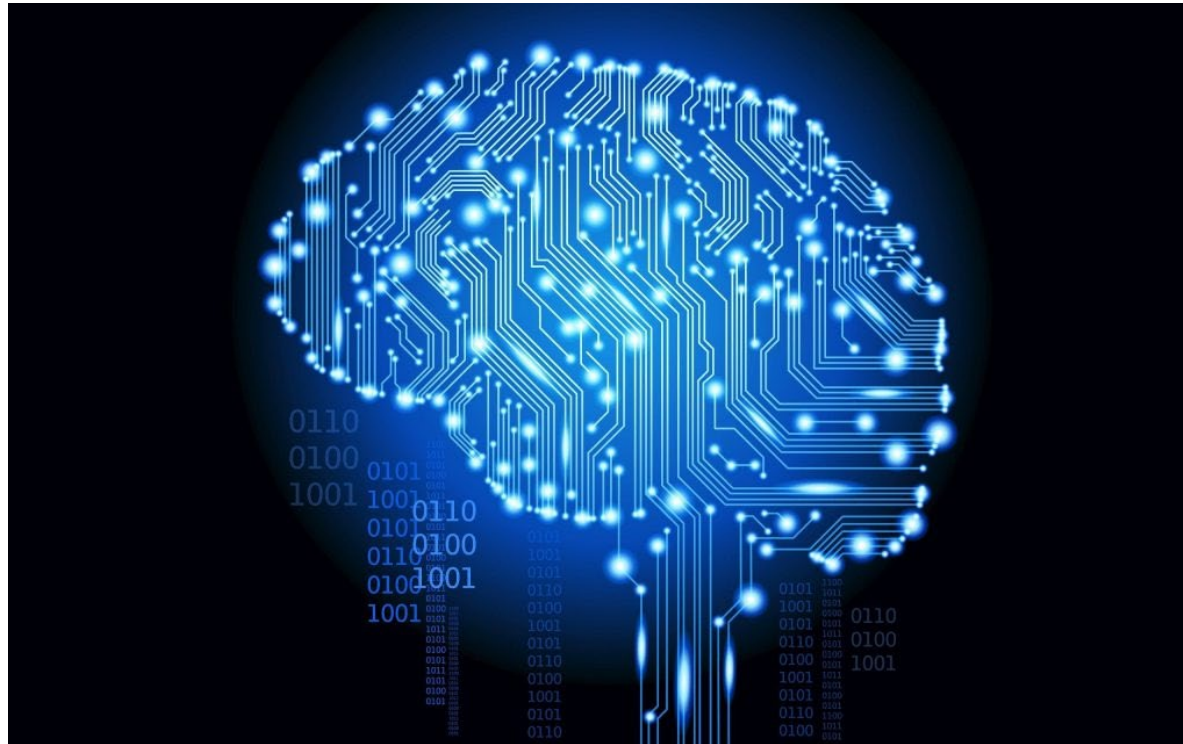


# Qu'est-ce que l'intelligence artificielle?

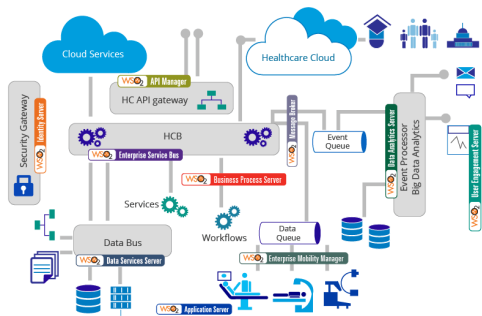
L'IA est une méthode scientifique qui confie à des machines des tâches exigeant de l'**intelligence** si elles étaient accomplies par des humains. Il s'agit d'un terme générique qui englobe :

- L'apprentissage machine
- Le traitement du langage naturel
- L'apprentissage profond
- Le traitement d'image
- La robotique
- D'autres tâches

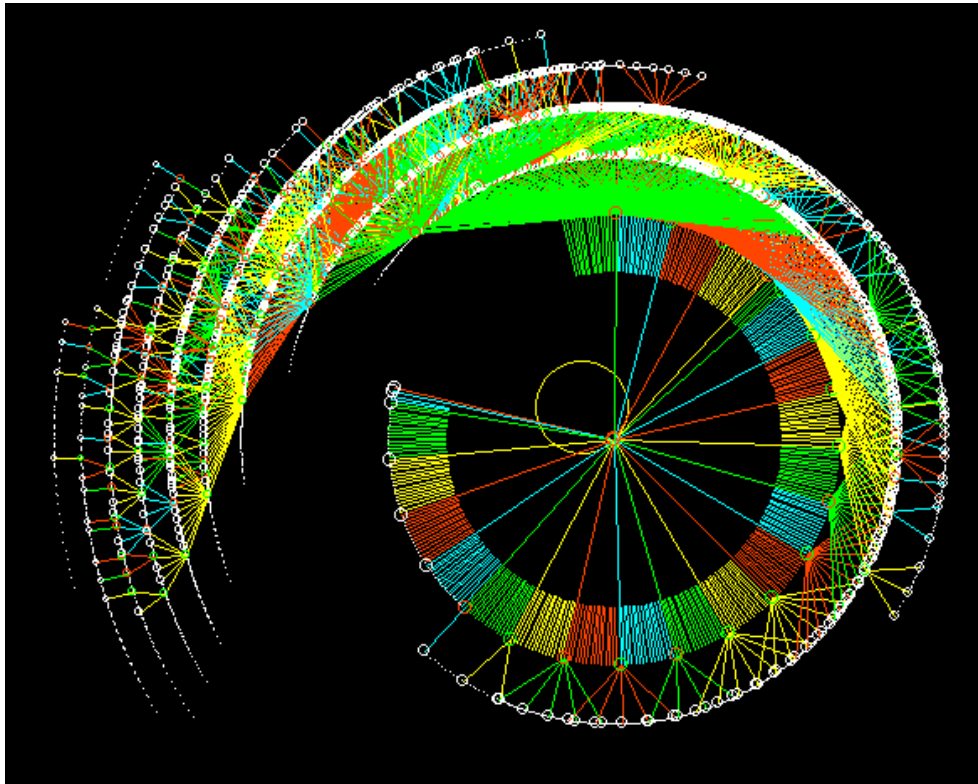
# Comment expliquer l'intérêt de plus en plus marqué envers l'intelligence artificielle?



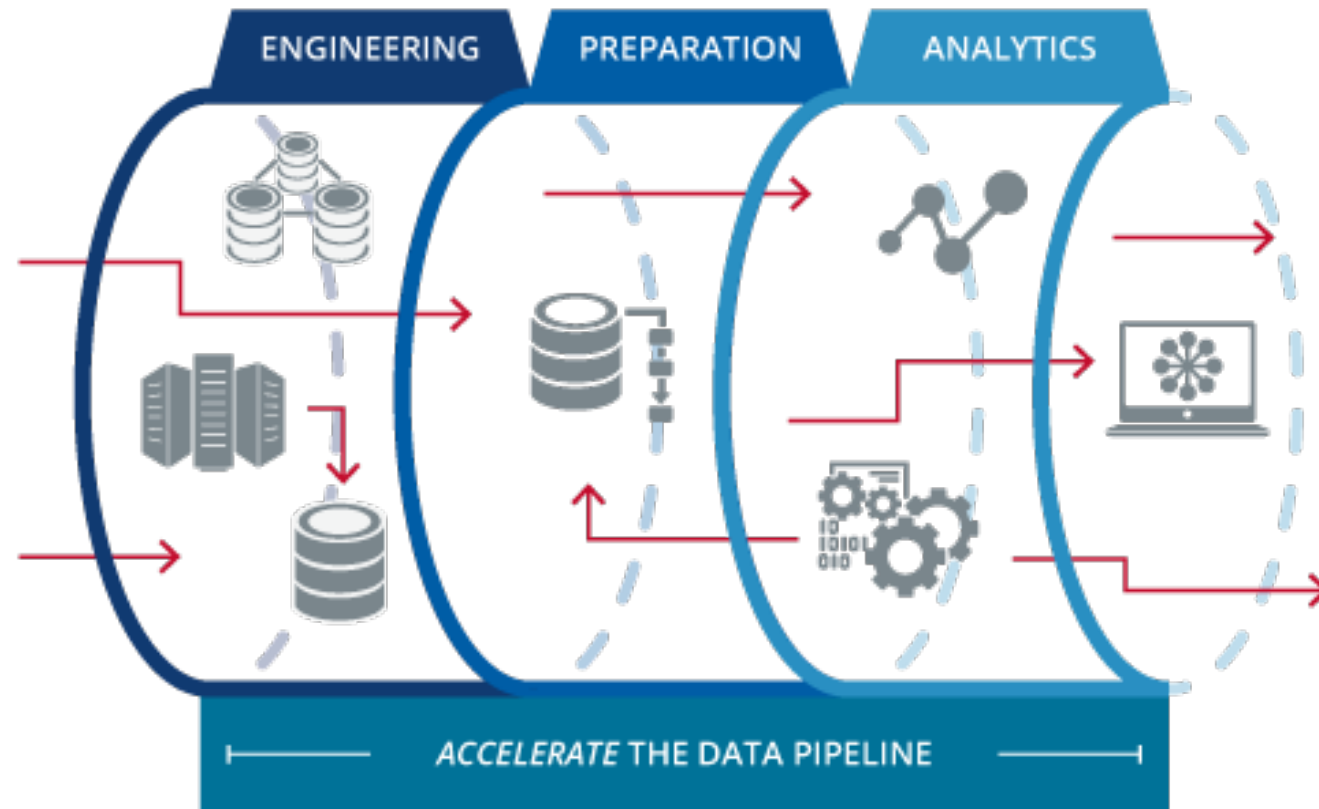
# Évolution rapide des systèmes de données



# Capacité informatique sans cesse croissante



# Améliorations apportées à l'ingestion et au traitement des données



# Besoin accru de décisions guidées par les données



DATA



KNOWLEDGE



ACTION

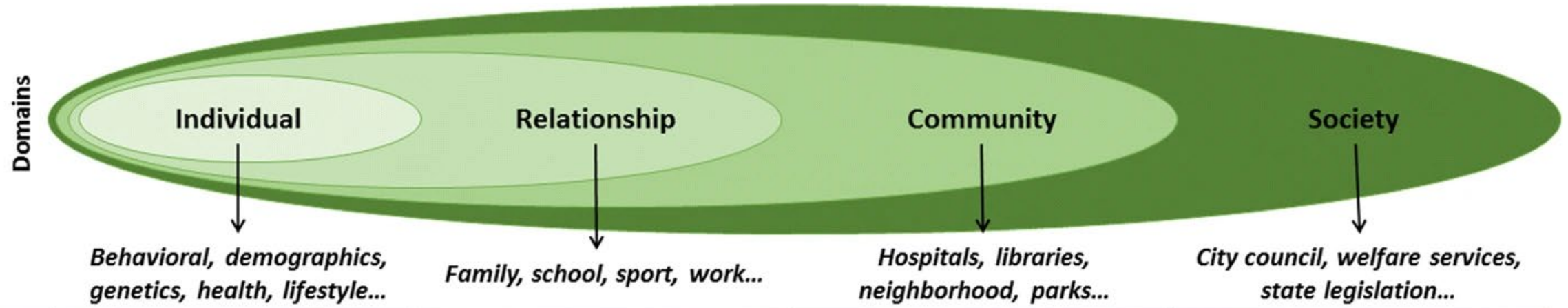


# CLARTÉ CONCEPTUELLE DES APPLICATIONS



- Expliquer le modèle
- Opérationnaliser le modèle





<b>Data Types</b>	Electronic health records Mobile apps Shopping receipts Social media posts Questionnaires Wearable techs	School records Social media networks Employment records	Area deprivation index, Crime rates, Food deserts, Green areas, Income, Pollution levels, Walkability	Gun control, Social security, Unemployment wage, Universal healthcare
<b>Example Sources</b>	Agency for Healthcare Research and Quality, National Health and Nutrition Examination Survey	Facebook, Instagram, National Center for Education Statistics, Twitter	American Community Survey, Department of Housing and Urban Development, Esri Demographics, Food Access Atlas, National Center for Environmental Protection, Terra Modis	CDC, CNN, FOX, White House

Qu'est-ce que  
l'apprentissage  
machine?

L'apprentissage machine est une  
branche de l'informatique qui  
permet à des ordinateurs  
d'**apprendre** sans être directement  
programmés pour le faire.

Terme désignant des techniques qui  
visent à ajuster des modèles  
algorithmiquement par adaptation  
aux schémas de données

L'apprentissage machine offre aux épidémiologistes de nouveaux outils pour s'attaquer à des problèmes difficiles à résoudre à l'aide des méthodes classiques.

---

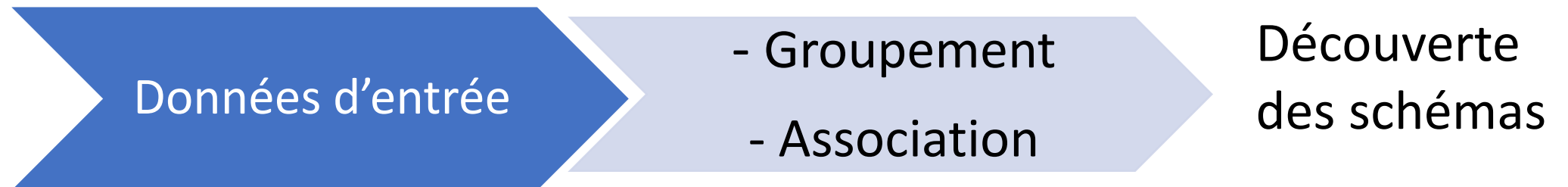
<b>Terme d'informatique</b>	<b>Terme d'épidémiologie</b>
Fonctionnalités	Variables indépendantes
Étiquettes	Résultats
Bruit des étiquettes	Erreur de mesure
Apprentissage	Ajustement
Exploration de données	Analyses exploratoires
Algorithme de classification	Algorithme avec variable-réponse catégorielle
Algorithme de régression	Algorithme avec variable-réponse continue
Précision	Valeur prédictive positive
Rappel	Sensibilité
Dimensionnalité	Nombre de covariables
Données déséquilibrées	Distribution inégale des résultats
Encodage one-hot	Création de variables nominales

---

## Apprentissage supervisé



## Apprentissage non supervisé



## Apprentissage semi-supervisé



**Table 3 Selected machine-learning approaches that have been applied to big data in public health**

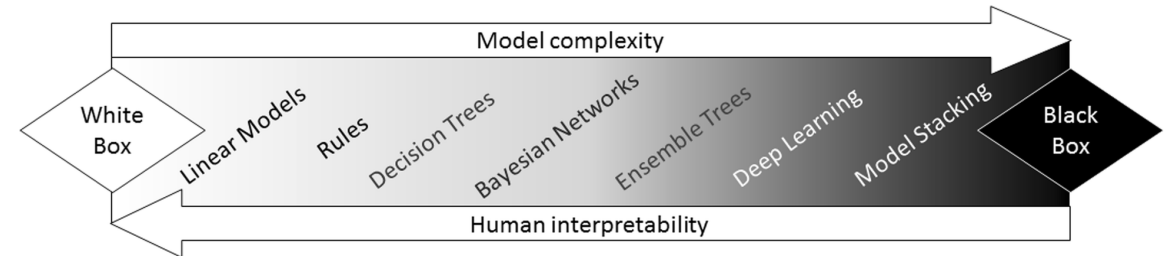
Approach	Learning type	Usage examples
<b>K-means clustering</b>	Unsupervised	Hot spot detection (4)
<b>Retrospective event detection</b>	Unsupervised	Case ascertainment (34)
<b>Content analysis</b>	Unsupervised	Public health surveillance (38)
<b>K-nearest neighbors clustering</b>	Supervised	Spatiotemporal hot spot detection (132); Clinical outcomes from genetic data; falls from wearable sensors
<b>Naïve Bayes</b>	Supervised	Acute gastrointestinal syndrome surveillance (51)
<b>Neural networks</b>	Supervised	Identifying microcalcification clusters in digital mammograms (100); predicting mortality in head trauma patients (31); predicting influenza vaccination outcome (126)
<b>Support vector machines</b>	Supervised	Diagnosis of diabetes mellitus (11); detection of depression through Twitter posts (27)
<b>Decision trees</b>	Supervised	Identifying infants at high risk for serious bacterial infections (8); comparing cost-effectiveness of different influenza treatments (115); and physical activity from wearable sensors (101)

# POSSIBILITÉS

- **Identification** plus *rapide* des **menaces émergentes** (p. ex., COVID-19)
- Compréhension plus *approfondie* et *actuelle* de la distribution des maladies et des facteurs de risque au sein de la population (p. ex., outils de surveillance des maladies en ligne; inspections ciblées de détection du plomb)
- **Prévision** de l'incidence d'une maladie dans le cadre de la planification de la santé de la population
- Meilleur ciblage des activités de promotion de la santé (p. ex., analyse de sentiments; outils et applications en ligne)
- Et plus encore (gestion de la santé de la population; effets de la modification des politiques; inférence causale)

# DÉFIS

- **Explicabilité**
- Biais
- *Accroissement* possible des **iniquités en matière de santé**
- Préoccupations concernant la **confidentialité**
- Accessibilité et partage des données
- Données et **infrastructure** analytique *désuètes*
- *Manque* de formation et de compétences sur l'IA au sein du milieu de la santé publique



**A health care algorithm affecting millions is biased against black patients**

*A startling example of algorithmic bias*

By Colin Lecher | @colinlecher | Oct 24, 2019, 2:00pm EDT



# Description

- Surveillance de la santé publique
- Études d'association

*« La surveillance de la santé publique désigne la collecte continue et systématique, l'analyse et l'interprétation de données de santé essentielles pour la planification, la mise en place et l'évaluation des pratiques en santé publique. »\**

# New Directions in Artificial Intelligence for Public Health Surveillance

Daniel B. Neill, *Event and Pattern Detection Laboratory, H.J. Heinz III College,  
Carnegie Mellon University*

## Detecting previously unseen outbreaks with novel symptom patterns

Yandong Liu and Daniel B. Neill\*

Event and Pattern Detection Laboratory, Carnegie Mellon University, Pittsburgh, PA, USA

- Utilisation de textes libres tirés des dossiers des services d'urgence dans le but de détecter, de localiser et de caractériser les **éclosions nouvelles et émergentes** de maladies

# Détection spatiale de cas basée sur texte libre

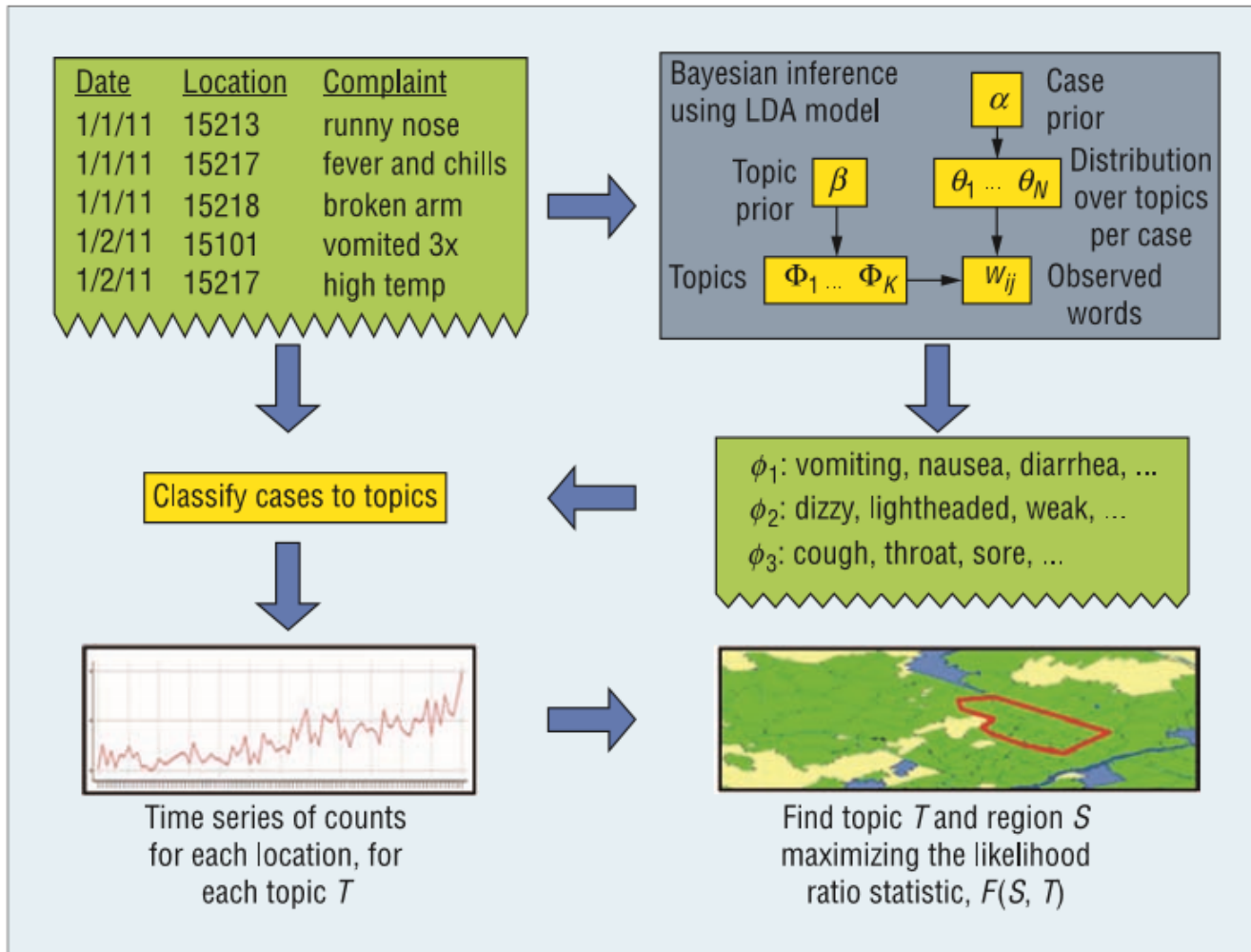


Figure 1. The semantic scan statistic learns a set of topics from the data using Latent Dirichlet Allocation, classifies each case into the most likely topic(s), and then maximizes a likelihood ratio statistic  $F(S, T)$  over all topics  $T$  and all space-time regions  $S$ .

<http://pophealthanalytics.com/>

- Les visites aux services d'urgence sont normalement regroupées dans des catégories : « maladie respiratoire », « maladie gastro-intestinale », « syndrome grippal, etc.
- Ce système définit l'objet de la requête selon les données.
- Par exemple, *cracher du sang* serait habituellement catégorisé en tant que « maladie respiratoire », ce qui **risque d'affaiblir le signal d'une éclosion et de retarder la détection** s'il y a de nombreux cas semblables.



# Real-time processing of social media with SENTINEL: A syndromic surveillance system incorporating deep learning for health classification

Ovidiu Șerban<sup>\*,1,a</sup>, Nicholas Thapen<sup>\*,1,a</sup>, Brendan Maginnis<sup>a</sup>, Chris Hankin<sup>a</sup>,  
Virginia Foot<sup>b</sup>

<sup>a</sup> *Institute for Security Science and Technology, Imperial College London, South Kensington Campus, London SW7 2AZ, UK*

<sup>b</sup> *The Defence Science and Technology Laboratory (DSTL), Porton Down, Salisbury SP4 0JQ, UK*

- Surveillance syndromique de l'actualité, des données cliniques et des données des médias sociaux à l'aide de réseaux de neurones profonds
- Traitement en temps réel de 1,8 million de gazouillis et de 18 000 articles de presse par jour
- Traitement du langage naturel (TLN), réseaux de neurones profonds, LASSO

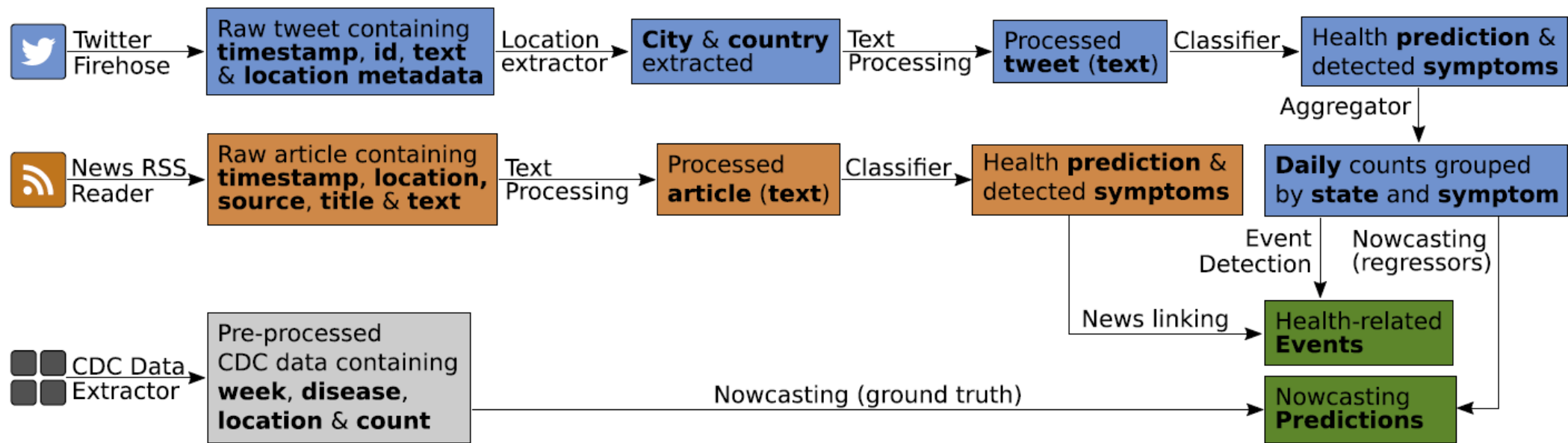


Fig. 1. A data integration diagram, showing the transformation process happening within SENTINEL.

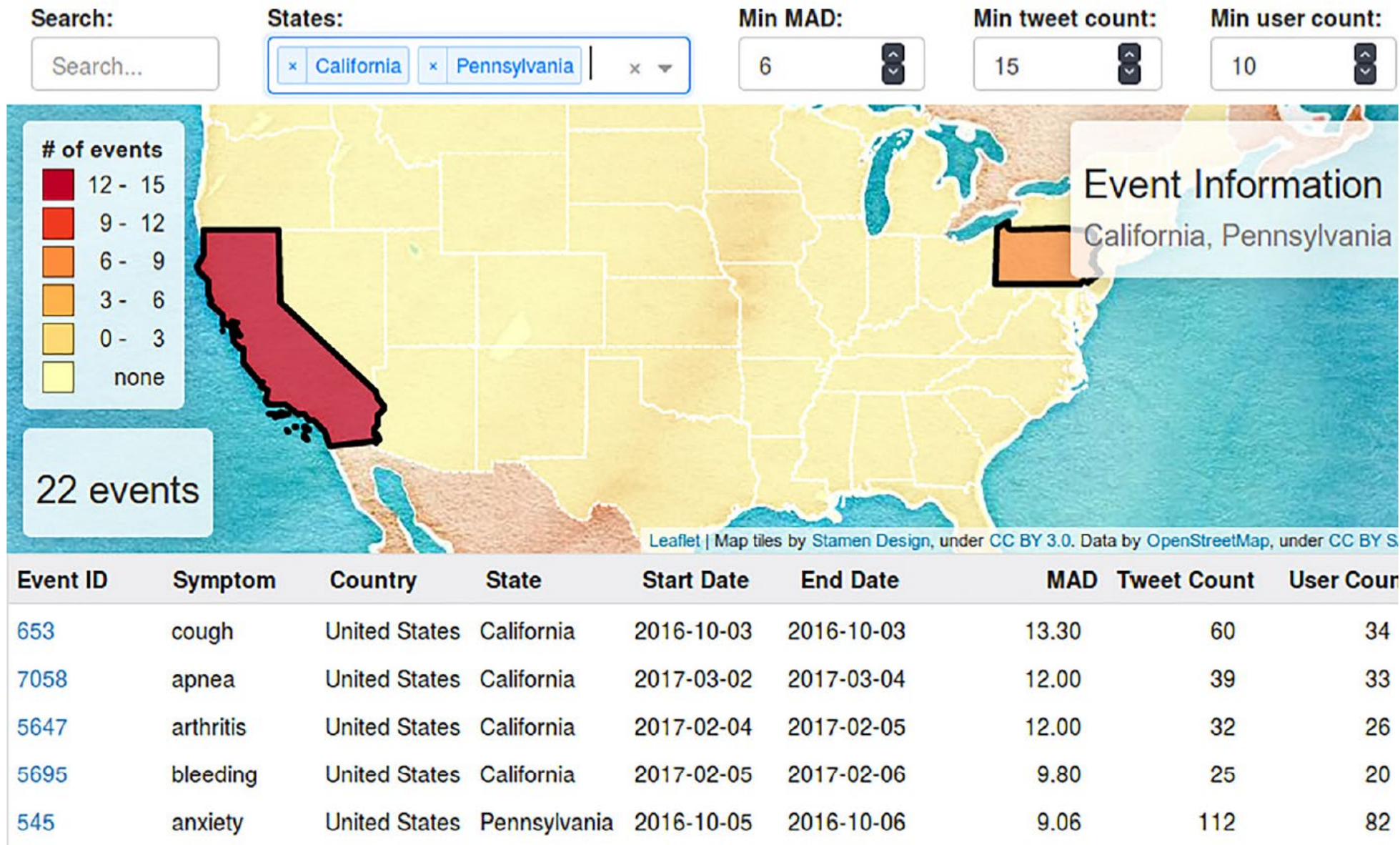


Fig. 3. The event list shown in the system.



Event ID	Symptom	Country	State	Start Date	End Date	MAD	Tweet Count	User Count
4986	flu	United States	Pennsylvania	2017-01-22	2017-01-24	7.70	44	27

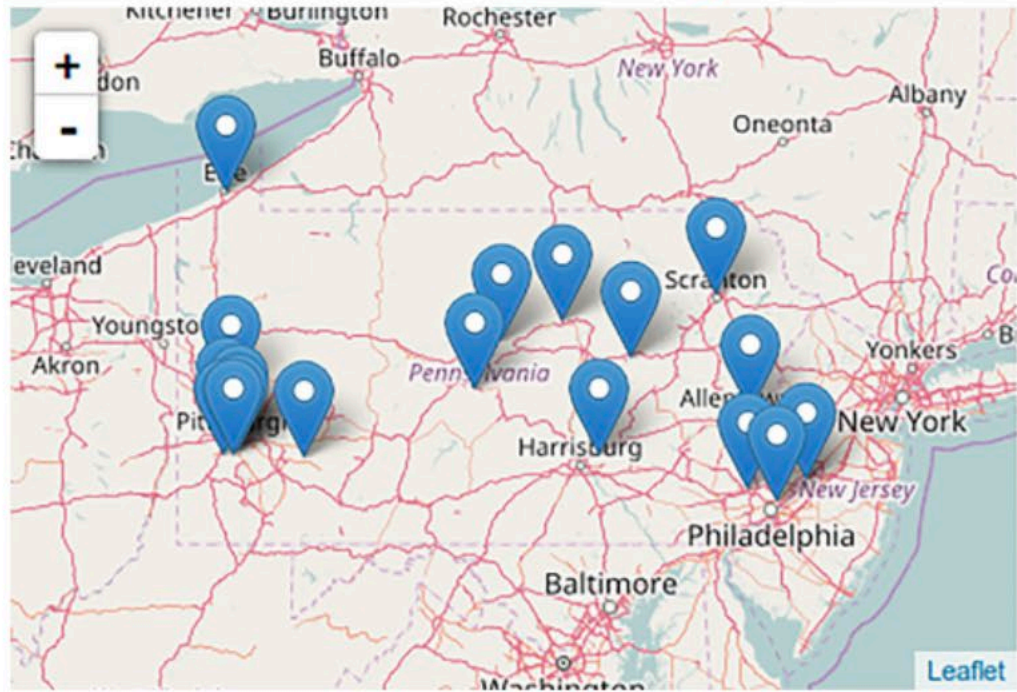
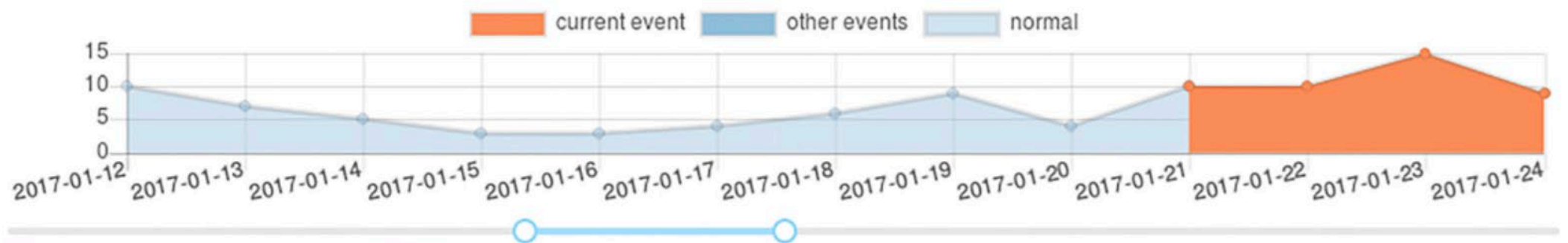
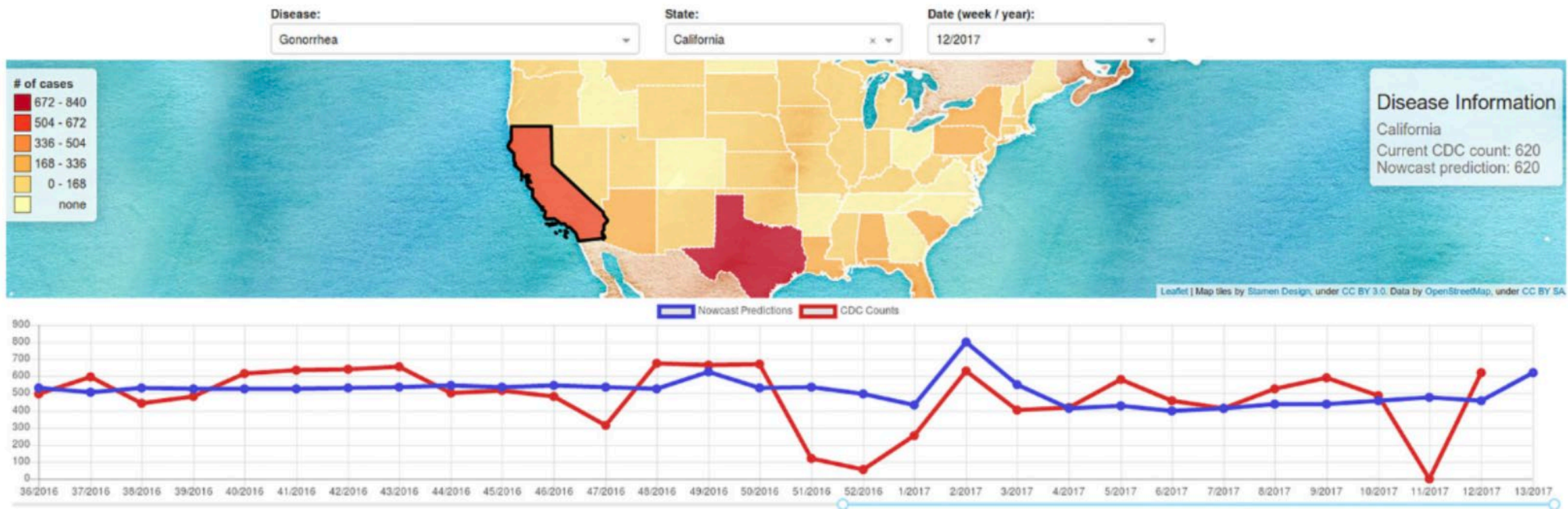


Fig. 4. The Situational Awareness page in the system - top half.



**Fig. 6.** The Nowcasting screen showing predictions for Gonorrhoea, in California on week 12 (2017).



# Prévision

- Géolocalisation entrée-sortie guidée par les données; ne peut déceler de mécanisme ni de causalité
- Diagnostic de la maladie vs incidence de la maladie; prévision individuelle vs prévision à l'échelle de la population
- Préoccupation concernant le surajustement
- Nécessité d'évaluer rigoureusement le modèle avant le déploiement





## Validation of a Machine Learning Model to Predict Childhood Lead Poisoning

Eric Potash, PhD; Rayid Ghani, MS; Joe Walsh, PhD; Emile Jorgensen, MPH; Cortland Lohff, MD, MPH; Nik Prachand, MPH; Raed Mansour, MS

Modèles de forêts aléatoires dérivés à partir de données recueillies entre 1997 et 2012 (N = 194 786); validation temporelle des données de 2013 (N = 6 182)

- A permis de prévoir l'exposition au plomb des enfants individuels et dans les domiciles
- Permet d'accorder la **priorité aux domiciles et à l'identification des enfants les plus à risque**
- Comparaison avec un modèle simple de régression logistique (facteurs prévisionnels incluant l'âge des membres du ménage, le revenu moyen, l'ethnicité, les externalités de voisinage fixes)

# DONNÉES

Modèle dérivé à partir de données recueillies entre 1997 et 2012 (N =194 786); validation temporelle des données de 2013 (N = 6 182)

- **Spatiales :**

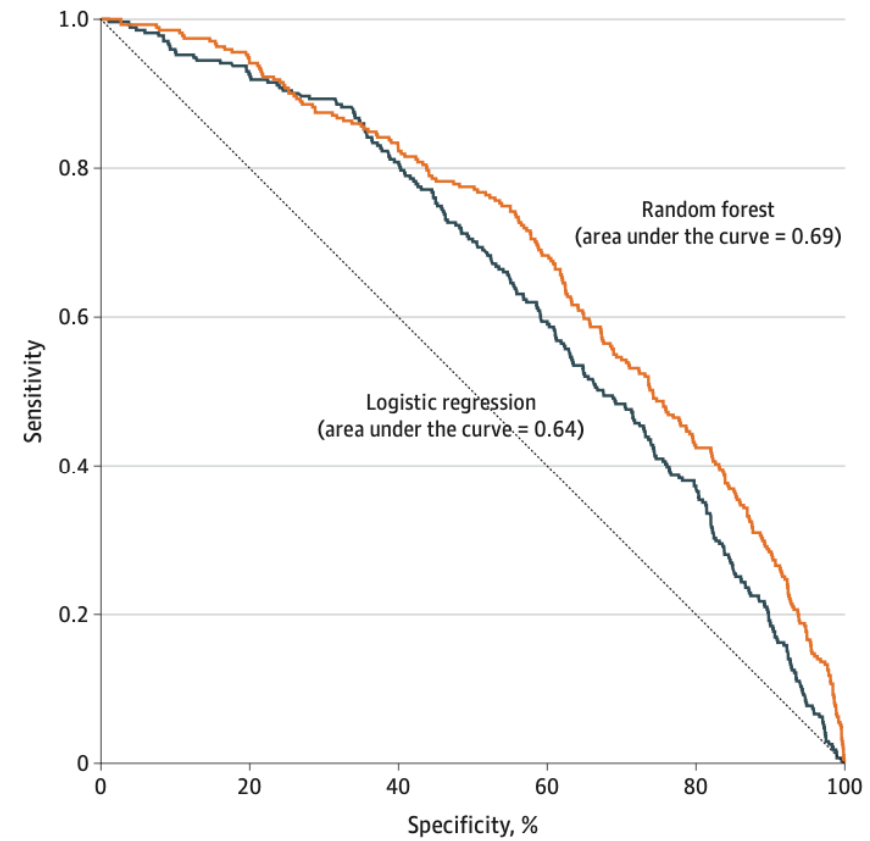
- Information sur l'immeuble incluant l'année de construction, la taille, l'état physique

- **Spatio-temporelles :**

- Résultats des analyses sur la concentration de plomb dans le sang (2,5 millions)
- Dossiers des inspections domiciliaires vérifiant la présence de plomb (70 000)
- Permis et infractions du Service du bâtiment de Chicago (2 millions)
- Variables sociodémographiques tirées du recensement incluant le niveau d'éducation, l'assurance-santé, les renseignements concernant la propriété immobilière

# RÉSULTATS

Figure 2. Receiver Operating Characteristic Curves for Random Forest and Logistic Regression Models



Difference in the areas under the receiver operating characteristics curve was 0.05 (95% CI, 0.02-0.08).

# RÉSULTATS

Figure 2. Receiver Operating Characteristic Curves for Random Forest and Logistic Regression Models

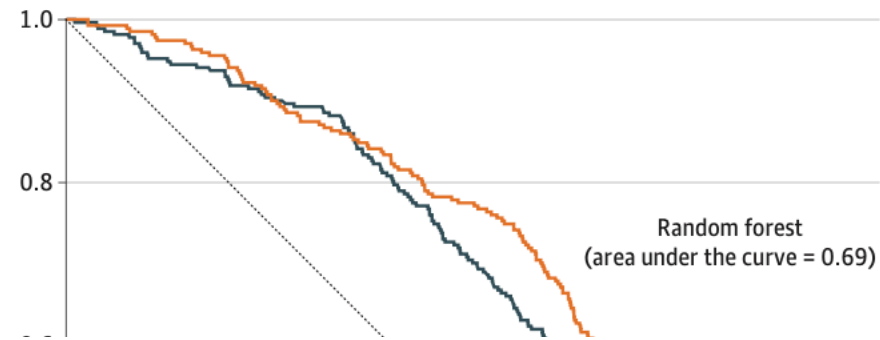
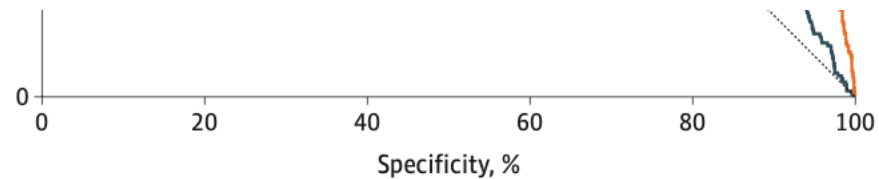


Table 2. Confusion Matrix Metrics for Random Forest and Logistic Regression Models

Population at highest risk, % <sup>a</sup>	Specificity, %			Sensitivity, %			PPV, %		
	Random forest	Logistic regression	Difference (95% CI) <sup>b</sup>	Random forest	Logistic regression	Difference (95% CI) <sup>b</sup>	Random forest	Logistic regression	Difference (95% CI) <sup>b</sup>
5	95.5	95.1	0.4 (0.0 to 0.7)	16.2	8.1	8.1 (3.9 to 11.7)	15.5	7.8	7.7 (3.7 to 11.3)
10	90.4	90.1	0.2 (-0.2 to 0.7)	27.3	19.9	7.4 (3.0 to 14.6)	12.7	9.4	3.3 (1.3 to 6.7)
20	80.3	79.9	0.3 (-0.1 to 1.4)	42.4	38.4	4.1 (-1.1 to 12.5)	9.9	8.9	1.0 (-0.1 to 3.0)



Difference in the areas under the receiver operating characteristics curve was 0.05 (95% CI, 0.02-0.08).

# RÉSULTATS

eTable 17. Confusion Matrix Metrics for the Random Forest Model by Race/Ethnicity

## Specificity

	All	Race/Ethnicity			
		Hispanic	Non-Hispanic Black	Non-Hispanic White	Asian
<b>Highest-risk %<sup>a</sup></b>					
5%	95.5%	97.1%	92.6%	97.4%	100.0%
10%	90.4%	93.3%	84.8%	97.0%	99.0%
20%	80.3%	84.3%	72.0%	94.0%	93.3%

## Sensitivity

	All	Race/Ethnicity			
		Hispanic	Non-Hispanic Black	Non-Hispanic White	Asian
<b>Highest-risk %<sup>a</sup></b>					
5%	16.2%	7.3%	26.0%	11.1%	0.0%
10%	27.3%	16.1%	38.9%	33.3%	0.0%
20%	42.4%	27.4%	58.8%	33.3%	14.3%

## Positive Predictive Value

	All	Race/Ethnicity			
		Hispanic	Non-Hispanic Black	Non-Hispanic White	Asian
<b>Highest-risk %<sup>a</sup></b>					
5%	15.5%	11.0%	17.5%	12.5%	NA
10%	12.7%	10.6%	13.5%	27.3%	0.0%
20%	9.9%	7.9%	11.3%	15.8%	4.5%

<sup>a</sup> Binary predictions are obtained from continuous risk scores by classifying this highest-risk percentage as positive.





# IMPLICATIONS

- ◎ Priorisation des inspections domiciliaires vérifiant la présence de plomb
- ◎ Publication des cotes de risque
- ◎ Intégration aux dossiers médicaux informatisés
- ◎ Sensibilisation des locateurs

**ARTICLE**    **OPEN**

# Machine-learned epidemiology: real-time detection of foodborne illness at scale

Adam Sadilek<sup>1</sup>, Stephanie Caty<sup>2</sup>, Lauren DiPrete<sup>3</sup>, Raed Mansour <sup>4</sup>, Tom Schenk Jr <sup>5</sup>, Mark Bergtholdt<sup>3</sup>, Ashish Jha<sup>2,6</sup>, Prem Ramaswami<sup>1</sup> and Evgeniy Gabrilovich<sup>1</sup>

- Foodborne IllNess DEtector in Real time (détecteur de maladies d'origine alimentaire en temps réel), aussi connu sous le nom de **FINDER** : modèle d'apprentissage machine qui permet de détecter les maladies d'origine alimentaire **en temps réel** à l'aide d'informations *anonymisées* et *agrégées* de recherche sur le Web et de données de localisation

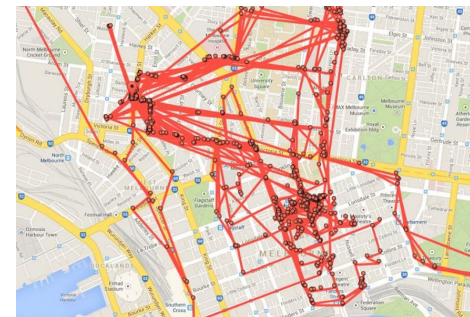
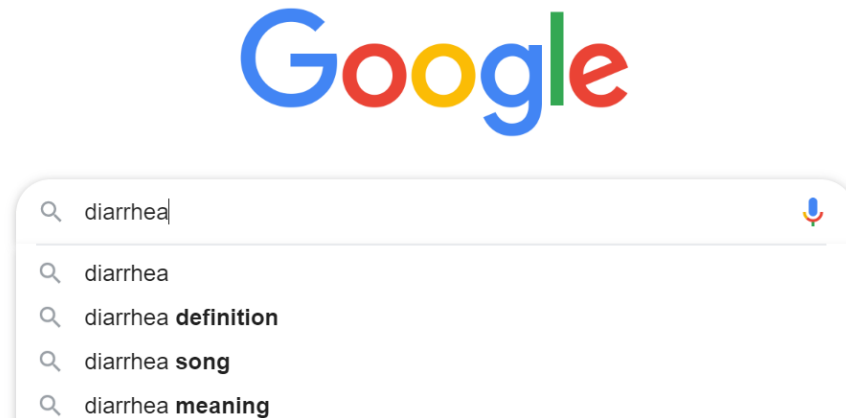
# DONNÉES

## Données anonymisées de recherche sur le Web

- À l'aide du traitement du langage naturel et des *pages Web sur les maladies d'origine alimentaire*, un modèle a été mis au point dans le but d'identifier les recherches en ligne sur les maladies d'origine alimentaire; un pointage entre 0 et 1 (c-statistique = 0,85) était attribué à ces recherches.

## Données anonymisées de localisation

- *Collecte de données de localisation ambiantes* trois jours avant une recherche en ligne sur les maladies d'origine alimentaire; elles sont recueillies automatiquement par Google lorsqu'il y a partage de localisation sur un appareil mobile.





# APPROCHE

- Les données de la recherche sur le Web sont utilisées pour évaluer la probabilité qu'une personne présente des symptômes de maladie d'origine alimentaire.
- Les données sur la localisation sont utilisées pour évaluer la **proportion de personnes ayant visité un restaurant** et effectuer par la suite une **recherche en ligne sur les maladies d'origine alimentaire**. Une liste des restaurants à risque élevé est ainsi produite.
- FINDER a été déployé dans deux circonscriptions sanitaires des États-Unis, à **Chicago** et **Las Vegas**, entre mai et août 2016.

# RÉSULTATS

**Table 1.** Number of inspections conducted during the experimental time period

	<b>FINDER</b>	<b>BASELINE</b>
Total	132	10,786
Las Vegas	61	4977
Chicago	71	5809
Complaint-driven	N/A	1291
Routine	N/A	4518
Risk level <sup>a</sup>		
High (% of total)	84 (63.6%)	5702 (52.9%)
Medium (%)	39 (29.6%)	2325 (21.6%)
Low (%)	9 (6.8%)	2759 (25.6%)

<sup>a</sup>*p* value for difference in risk distribution between FINDER and BASELINE <0.001, from  $\chi^2$ -test

# RÉSULTATS

**Table 2.** Ability of FINDER to detect unsafe restaurants as compared to BASELINE rate and with subcategories of the baseline inspections, including complaint-based inspections that occurred in Chicago and routine inspections from both Chicago and Las Vegas

	FINDER <i>n</i> = 132	BASELINE <i>n</i> = 10,786	Odds ratio <sup>a</sup> [95% CI]	<i>p</i> -value
Overall, number unsafe (%)	69 (52.3%)	2662 (24.7%)	3.06 [2.14–4.35]	<0.001
Risk level				
High, number unsafe (%)	42 (50.0%)	1909 (33.5%)	1.98 [1.28–3.05]	0.002
Medium, number unsafe (%)	23 (59.0%)	536 (23.1%)	5.50 [2.83–10.72]	<0.001
Low, number unsafe (%)	4 (44.4%)	217 (7.9%)	7.35 [1.79–30.13]	0.006
Comparison of FINDER to complaint-based inspections				
	<b>FINDER</b> <i>n</i> = 71	<b>COMPLAINT</b> <i>n</i> = 1291		
Overall, number unsafe (%)	37 (52.1%)	508 (39.4%)	1.68 [1.04–2.71]	0.03
Risk level				
High, number unsafe (%)	27 (47.4%)	374 (39.4%)	1.38 [0.81–2.36]	0.24
Medium, number unsafe (%)	9 (75.0%)	115 (39.3%)	4.64 [1.23–17.51]	0.02
Low, number unsafe (%)	1 (50.0%)	19 (38.8%)	1.58 [0.09–26.78]	0.75

# IMPLICATIONS

- ◎ Inspections plus efficaces des restaurants et réduction des maladies d'origine alimentaire
- ◎ Identification plus précise de ???
- ◎ Il faut toutefois prendre sérieusement en considération la petite taille des échantillons et les préoccupations sur le plan de la confidentialité des renseignements.



## Inférence causale

- Même si la plupart des techniques d'AI/AM sont utilisées à des fins descriptives et prévisionnelles, on peut se servir *des mêmes outils à des fins de causalité*.
- Les utilisations actuelles de l'IA/AM ne modifient pas nécessairement notre compréhension conceptuelle des paradigmes de causalité, mais fournissent un nouvel outil pour nous aider à explorer la causalité.

# The impact of social housing on mental health: longitudinal analyses using marginal structural models and machine learning-generated weights

Rebecca Bentley,<sup>1,2\*</sup> Emma Baker,<sup>2</sup> Koen Simons,<sup>3</sup> Julie A Simpson<sup>3†</sup>  
and Tony Blakely<sup>3,4†</sup>

- Examen des répercussions d'une **exposition cumulative au logement social et des transitions à un logement social sur la santé mentale** à l'aide des données d'un sondage sur la santé
- Utilisation de modèles structuraux marginaux avec pondération par probabilité inverse de traitement (Inverse Probability of Treatment Weights ou IPTW) dérivée à l'aide de **méthodes d'apprentissage ensembliste**
  - Ces méthodes combinent plusieurs modèles d'apprentissage machine.

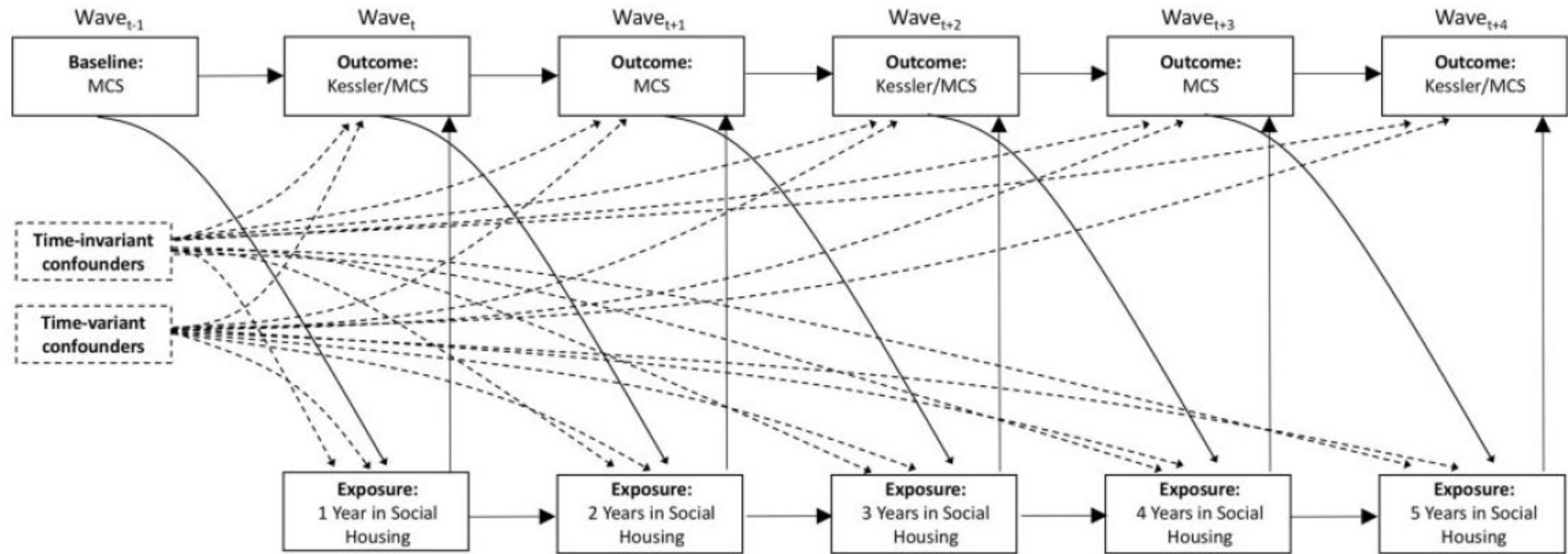


Figure 1. Directed acyclic graph (DAG).

- La pondération par probabilité inverse de traitement englobait un grand nombre de facteurs de confusion de base et variant dans le temps, ainsi qu'un score de santé mentale de l'année précédente.
- Les personnes constamment exposées au logement social avaient en moyenne un score de santé mentale inférieur à celui des sujets non exposés.

# 5 types de mégadonnées

Source	Examples	Aspect of bigness <sup>a</sup>	Key technical issues	Typical uses
<b>-omic/biological</b>	Whole exome profiling, metabolomics	Wide	Lab effects, informatics pipeline	Etiologic research, screening
<b>Geospatial</b>	Neighborhood characteristics	Wide	Spatial autocorrelation	Etiologic research, surveillance
<b>Electronic health records</b>	Records of all patients with hypertension	Tall, often also wide	Data cleaning, natural language	Clinical research, surveillance
<b>Personal monitoring</b>	Daily GPS records, Fitbit readings	Tall	Redundancy, inference of intentions	Etiologic research, potentially clinical decision making
<b>Ready-made data</b>	Google search results, Reddit	Tall	Selection biases, natural language	Surveillance, screening, identification of hidden social networks



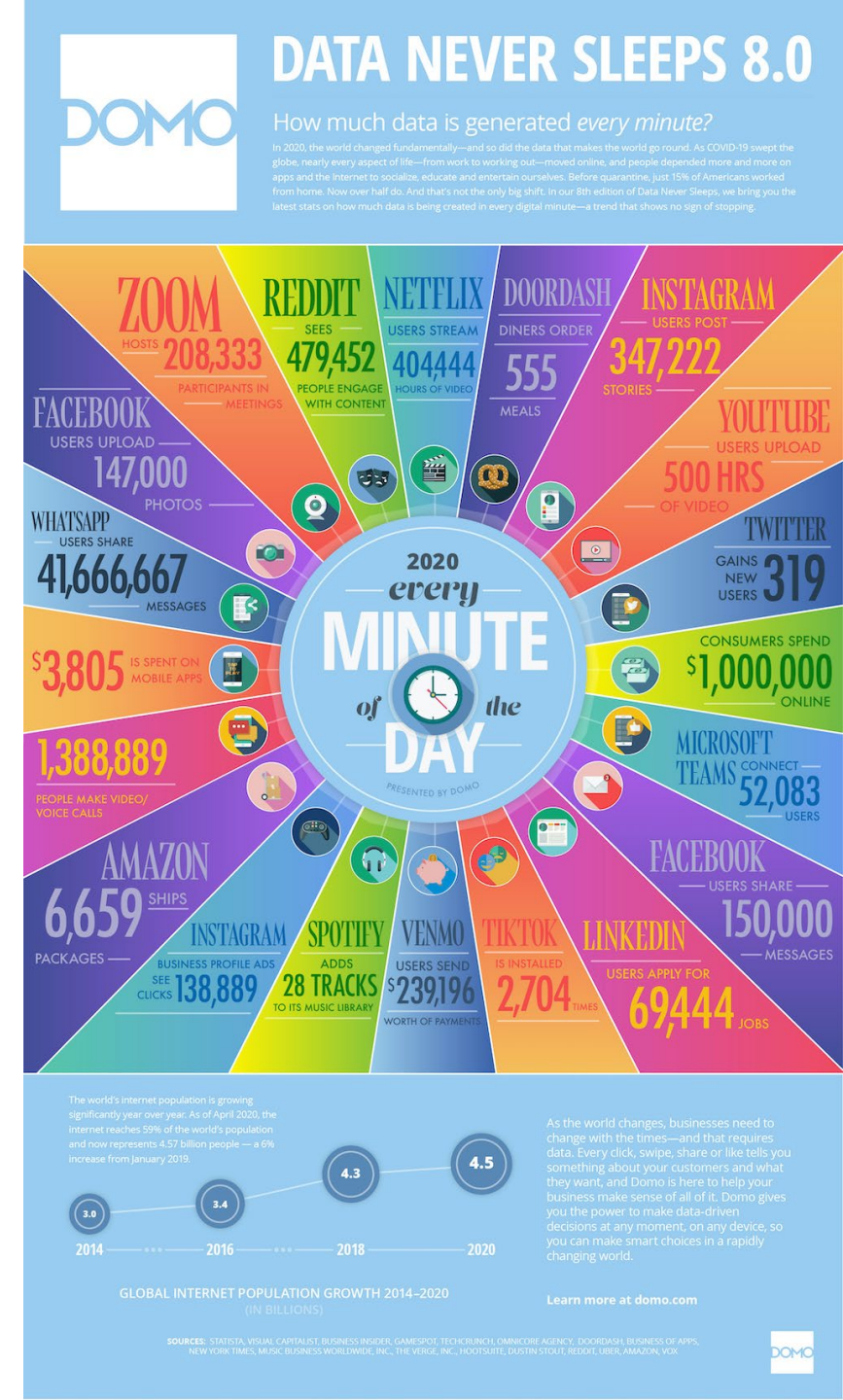
# Traces numériques prêtes à l'emploi

## Avantages

- Mégadonnées
- Toujours actives
- Non réactives
- Décrivent les relations sociales
- Faible coût





## Inconvénients

- Incomplètes
- Inaccessibles
- Non représentatives
- Non stationnaires
- Effet confusionnel sur les algorithmes
- Validité incertaine



# Application aux maladies non transmissibles

## Estimating geographic subjective well-being from Twitter: A comparison of dictionary and data-driven language methods

Kokil Jaidka<sup>a,b,1</sup> , Salvatore Giorgi<sup>c</sup>, H. Andrew Schwartz<sup>d</sup> , Margaret L. Kern<sup>e</sup> , Lyle H. Ungar<sup>c</sup>,  
and Johannes C. Eichstaedt<sup>f,g,1</sup> 

<sup>a</sup>Department of Communications and New Media, National University of Singapore, Singapore 117416; <sup>b</sup>Centre for Trusted Internet and Community, National University of Singapore, Singapore 117416; <sup>c</sup>Department of Computer and Information Science, University of Pennsylvania, Philadelphia, PA 19104; <sup>d</sup>Department of Computer Science, Stony Brook University, Stony Brook, NY 11794; <sup>e</sup>Melbourne Graduate School of Education, The University of Melbourne, Parkville, VIC 3010, Australia; <sup>f</sup>Department of Psychology, Stanford University, Stanford, CA 94305; and <sup>g</sup>Institute for Human-Centered Artificial Intelligence, Stanford University, Stanford, CA 94305

# Application aux maladies transmissibles

## Forecasting Zika Incidence in the 2016 Latin America Outbreak Combining Traditional Disease Surveillance with Search, Social Media, and News Report Data

**Sarah F. McGough<sup>1,2,3\*</sup>, John S. Brownstein<sup>2,3,4</sup>, Jared B. Hawkins<sup>2,3,4</sup>,  
Mauricio Santillana<sup>2,3,4\*</sup>**

**1** Harvard T.H. Chan School of Public Health, Boston, Massachusetts, United States of America, **2** Computational Health Informatics Program, Boston Children's Hospital, Boston, Massachusetts, United States of America, **3** Computational Epidemiology Group, Division of Emergency Medicine, Boston Children's Hospital, Boston, Massachusetts, United States of America, **4** Department of Pediatrics, Harvard Medical School, Boston, Massachusetts, United States of America





# Prévoir l'incidence de l'infection à virus Zika à l'aide d'une approche d'ensemble en AM dans les Amériques

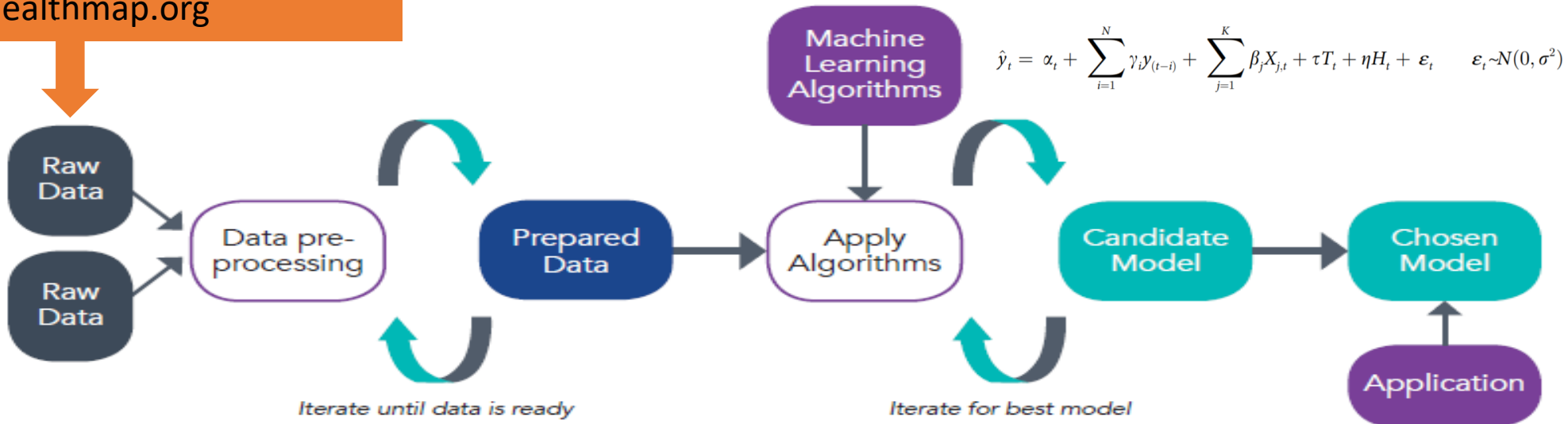
Données épidémiologiques :

- OPS
- Données nationales du ministère de la Santé

Traces numériques :

- Twitter
- Recherche à l'aide de Google Correlate et Google Trends
- Healthmap.org

Model	3 week		
	RMSE	rRMSE	$\rho$
AR	886.701	555.937	-0.903
G+T	<b>292.718</b>	<b>64.733</b>	<b>0.355</b>
ARGO+T	323.089	158.377	0.243
ARGO+TH	335.778	163.436	0.085



# Applications visant à atténuer les effets néfastes de l'infodémie

## Gestion de l'infodémie sur la COVID-19 : Promouvoir des comportements sains et atténuer les effets néfastes de la diffusion d'informations fausses et trompeuses

Déclaration conjointe de l'OMS, des Nations Unies, de l'UNICEF, du PNUD, de l'UNESCO, de l'ONUSIDA, de l'UIT, de l'initiative Global Pulse et de la FICR



台灣事實查核中心  
Taiwan FactCheck Center

**Détection de foules, collecte en temps réel, groupement automatisé**

I predict that the next major outbreak – whether of a highly fatal strain of influenza or something else – will not be due to a lack of preventive technologies. Instead, emotional contagion, digitally enabled, could erode trust in vaccines so much as to render them moot. The deluge of conflicting information, misinformation and manipulated information on social media should be recognized as a global public-health threat.

Larson, Heidi J. “The Biggest Pandemic Risk? Viral Misinformation.” *Nature* 562, no. 7727 (16 octobre 2018): 309–309.

# Possibilités d'utiliser des données prêtes à l'emploi en AI dans le domaine de la santé publique

## Possibilités :

- Ouvertes et accessibles
- Exécution rapide
- Interdisciplinaire

## Défis :

- Acquisition de données
- Consentement éclairé
- Confidentialité des renseignements
- Prendre des décisions en contexte d'incertitude

Dataset

## COVID-19 Open Research Dataset Challenge (CORD-19)

An AI challenge with AI2, CZI, MSR, Georgetown, NIH & The White House

 Allen Institute For AI and 8 collaborators • updated 2 days ago (Version 59)

Data [Tasks \(17\)](#) Notebooks (1,627) Discussion (370) Activity Metadata

New Notebook



# Récapitulation

Définir l'orientation de la santé publique



# Tension entre nos interventions auprès des individus et nos interventions auprès de la population

- « Il n'est pas nécessaire d'ajouter le mot "précision" au concept de santé publique. »
- L'accent est mis au mauvais endroit (pas sur les déterminants sociaux de la santé). (traduction libre)

-Chowkwanyun, Bayer, Galea, NEJM 2018

- « La santé publique de précision offre une possibilité remarquable de revigorer une discipline qui n'a jamais joué un rôle aussi important dans l'amélioration de la santé de nos communautés les plus vulnérables et marginalisées. » (traduction libre)
- -Horton, Lancet, 2018



Est-ce que l'IA et l'AM feront progresser la santé publique? Seulement si....

---

Un moins grand nombre de personnes sont atteintes de maladies évitables.

---

La promotion de la santé se fait dans nos environnements et communautés.

---

On parvient à réduire les iniquités en matière de santé.

# Résumé

- **Pour mettre l'accent sur la santé de la population, nous avons besoin de sources plus vastes pour obtenir des ensembles de données sur la population**
  - Équité en IA = approche de santé publique
- **Les questions et les priorités en matière de santé de la population sont cohérentes; les outils émergents et les nouvelles sources de données sont prometteurs.**
  - Il faut accroître les liens vers un plus vaste éventail de sources de données qui reflètent de manière plus juste les déterminants de la santé et les nouvelles façons de travailler avec ces données.
  - Nous devons envisager de plus en plus souvent des moyens d'intégrer des outils prospectifs de planification à l'échelle de la population.
  - Nous devons nous montrer plus stratégiques en prenant des mesures pour répondre aux besoins de santé de la population et accélérer le cycle données-action.
- **Les méthodes d'IA/AM peuvent avoir de multiples applications à différentes fins; il faut miser sur la clarté.**
  - Clarifier la nature du problème et la raison pour laquelle la méthode choisie convient le mieux aux données ou au problème.
  - Description/surveillance, prévision et inférence causale
- **Les problèmes liés aux biais exigent une attention de tous les instants.**
  - Erreur de mesure/qualité des données
  - Biais de sélection

# Question de sondage

Si vous étiez aux commandes, que feriez-vous pour favoriser l'utilisation de l'IA/AM dans votre milieu?

# Ressources éducatives

- Cours intensif (anglais) sur l'intelligence artificielle ([youtube.com/crashcourse](https://www.youtube.com/watch?v=crashcourse))
- Éléments de l'IA (anglais) (<http://course.elementsofai.com>)
- Coursera : Apprentissage machine (en anglais, Université Stanford) (<https://www.coursera.org/learn/machine-learning>)
- Trousse de recherche sur les médias sociaux (en anglais, Université Ryerson) (<http://socialmediadata.org/social-media-research-toolkit/>)

# Merci!

**Laura C. Rosella, Ph.D.**

[laura.rosella@utoronto.ca](mailto:laura.rosella@utoronto.ca)

@LauraCRosella

**Stacey Fisher, Ph.D.**

[stacey.fisher@utoronto.ca](mailto:stacey.fisher@utoronto.ca)

@StaceyFisher\_

**Melodie Song, Ph.D.**

[melodie.song@oahpp.ca](mailto:melodie.song@oahpp.ca)

@MelodieYJSong

Public  
Health  
Ontario

Santé  
publique  
Ontario



Population Health  
Analytics Laboratory

**Dalla Lana**  
School of Public Health