



Machine Learning

Responses to Covid-19



Covid-19 ML Focus Areas

Reporting

- Regional and global spread
- Effects of social distancing and quarantine
- Assessment of government measures

Medical Research

- Antibody identification
- Protein, pathogen and genome sequencing
- CT image analysis
- Treatment and drug analysis

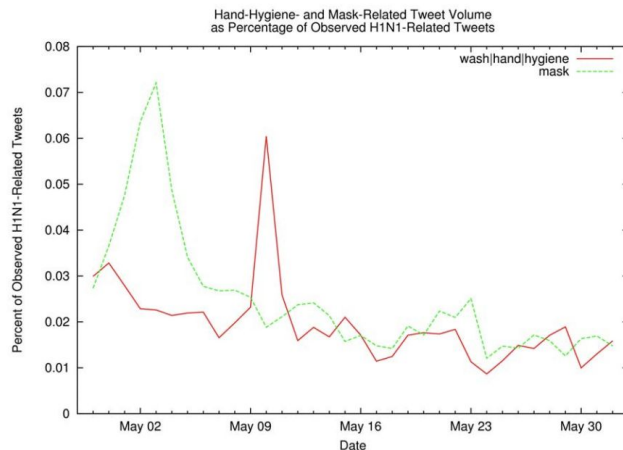
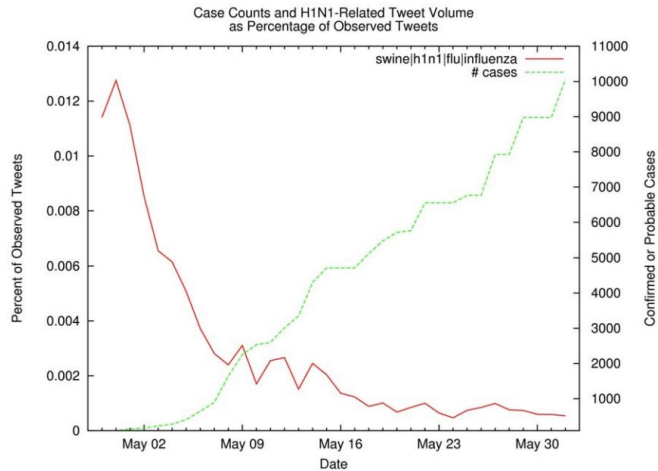
Risk Assessment

- Forecast modeling
- Predicting infection and patient risk
- Biometric analysis
- Contact tracing

Reporting:

Twitter to Track Levels of Disease Activity and Public Concern (Iowa)

- University of Iowa examined US-based tweets trained on Influenza-like illness (ILI) language using the Twitter API
- Dictionary built w/ tweets containing flu, swine, influenza, vaccine, etc., cleaned hashtags, etc. & applied Porter Stemming Method for **support-vector regression model**
- Estimation model built with open-source libSVM implementation trained weekly term-frequency statistics, using the ILI values for weeks in Oct 2009 - May 2010
- Twitter language can be used descriptively to track users' interest and concerns related to H1N1 and to **estimate disease activity 1–2 weeks ahead of other practices**



Reporting:

COVID-19 Epidemic Analysis using Machine Learning & Deep Learning Algorithms (India)

- Paper on medRxiv from 4 scientists modelled data from Johns Hopkins predicts the possible number of cases for the next 10 days around the world
- Regression approaches trained/tested on realtime data using the # of confirmed, recovered, & death cases as the label for the day; cases were scaled using minmax transformation from sklearn
- Complexly layered network: Deep neural network (DNN) dense input layer, a single Recurrent Neural Networks (RNN) output layer of 3 LSTM stacks + 10% dropout to avoid overfitting and final output layer with a single neuron
- Polynomial regression is best fit (min. root mean sq. error)

Table 1 Comparative results.

Model	RMSE			10-04-2020 (Forecast)		
	Confirmed	Deaths	Recovered	Confirmed	Deaths	Recovered
SVR	27456.47	1360.47	16762.15	1970860	93112	422740
DNN	163335.65	8554.55	25415.03	211213	7419	93476
LSTM	15647.64	1076.06	4092.01	1002543	50362	206916
PR	455.92	117.94	809.71	2156540	124021	557125

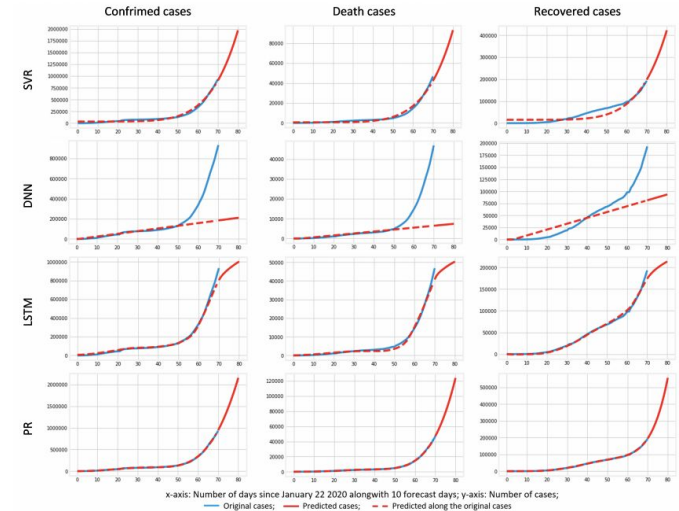


Fig. 6 COVID-19 worldwide epidemic analysis using SVR, DNN, LSTM, and PR.

Additional reporting projects

Prediction

1. Were 21% of New York City residents really infected with the novel coronavirus? (ML- adjacent, statistical bias assessment): [Towards Data Science](#)
2. AI-based model HealthMap, at Boston Children's Hospital in the USA sounded an alarm on 30 December 2019: www.diseasedaily.org/about
3. BlueDot predicted the outbreak of the infection at the end of 2019, issuing a warning 31 December 2019: bluedot.global/

Quarantine and social distancing

1. Social distancing detection for COVID-19: mc.ai/social-distancing-detection-for-covid-19/
2. Tracking government responses towards COVID-19: coronanet-project.org/
3. Quantifying the effect of quarantine control in Covid-19 infectious spread using machine learning: www.medrxiv.org/content/10.1101/2020.04.03.20052084v1
4. Robert Koch Institute in Berlin uses an epidemiological SIR model that takes into account containment measures by governments, such as lockdowns, quarantines, and social distancing prescriptions: rocs.hu-berlin.de/corona/docs/forecast/model/

Medical Research:

Computational predictions of protein structures associated with COVID-19 (DeepMind)

- AlphaFold deep learning system uses “free modeling” predicts protein structure when no structures of similar proteins are available
- Trained a neural network to predict a distribution of distances between every pair of residues in a protein
- Probabilities combined into a score that estimates how accurate a proposed protein structure is and scoring functions were used to find structures that matched
- Optimised scores with gradient descent
- Confirmed model provided an accurate prediction for the experimentally determined SARS-CoV-2 spike protein structure shared in the [Protein Data Bank](#)

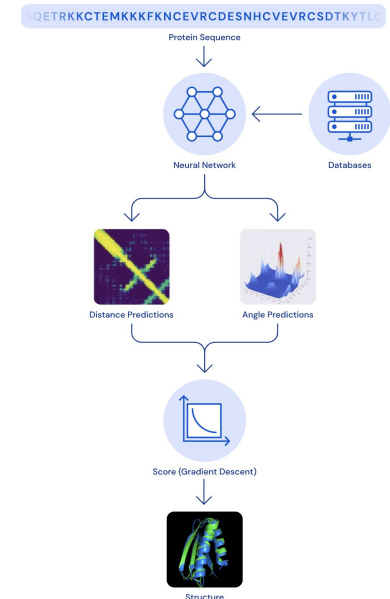
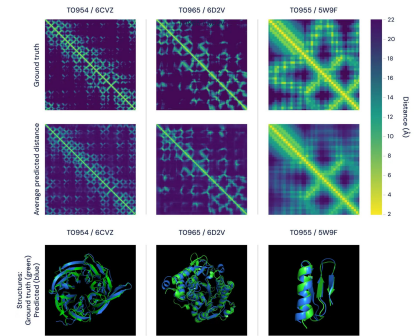


FIGURE 3. A SCHEMATIC OF THE ARCHITECTURE OF THE ALPHAFOLD SYSTEM PREDICTING STRUCTURE FROM PROTEIN SEQUENCE.

Medical Research:

Lung Infection Quantification of COVID-19 in CT images with Deep Learning (Cornell)

- Deep learning segmentation system is developed to automatically quantify infection regions in proportion to the lung
- Human-in-the-loop (radiologists, etc.) for faster identification of training samples and manual intervention of automated results
- Performance evaluated by comparing automatically segmented infection regions with the manually-delineated ones on 300 chest CT scans of 300 COVID-19 patients
- 91.6% agreement between automatic and manual infection segmentations

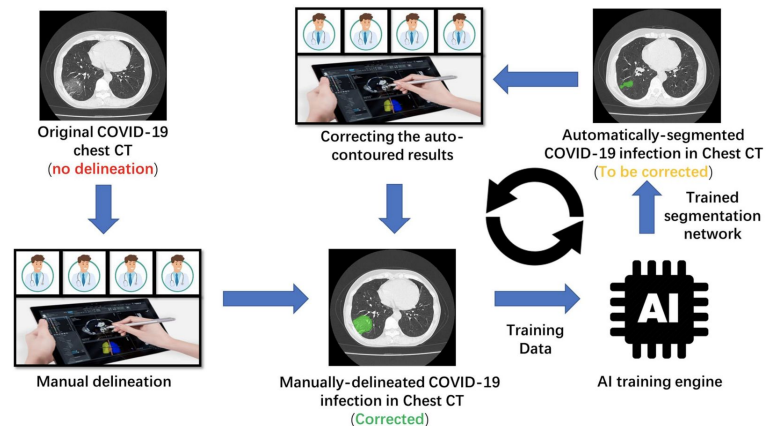


Figure 2. The human-in-the-loop workflow

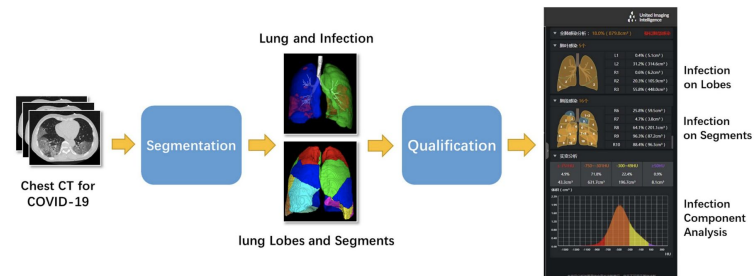


Figure 3 Pipeline for quantifying COVID-19 infection. A chest CT scan is first fed into the DL-based segmentation system. Then, quantitative metrics are calculated to characterize infection regions in the CT scan, including but not limited to infection volumes and POIs in the whole lung, lung lobes and bronchopulmonary segments.

Additional medical research projects

Understanding Covid19

1. Machine learning using intrinsic genomic signatures for rapid classification of novel pathogens: COVID-19 case study: journals.plos.org/plosone/article?id=10.1371/journal.pone.0232391
2. Virtual screening approach to identifying influenza virus neuraminidase inhibitors using molecular docking combined with machine-learning-based scoring function: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5669956/>
3. Over 15,000 viral genome sequences of hCoV-19 shared with unprecedented speed via GISAID: www.gisaid.org/

Diagnosis and prognosis

4. COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest X-Ray Images: arxiv.org/abs/2003.09871
5. CAD4COVID, is “an artificial intelligence software that triages COVID-19 suspects on chest X-rays images”: www.delft.care/cad4covid/
6. ML to develop a prognostic prediction algorithm to predict the likelihood of someone surviving the infection: www.ttopstart.com/news/how-ai-can-help-fight-covid-19
7. Rapid AI Development Cycle for the Coronavirus (COVID-19) Pandemic: Initial Results for Automated Detection & Patient Monitoring using Deep Learning CT Image Analysis: arxiv.org/abs/2003.05037v1

Additional medical research projects

Treatments

1. DSI Alumni Use Machine Learning to Discover Coronavirus Treatments:
datascience.columbia.edu/dsi-alumni-use-machine-learning-discover-coronavirus-treatments
2. Neural integration of neighbor information from a heterogeneous network for discovering new drug-target interactions:
www.researchgate.net/publication/330075138_NeoDTI_neural_integration_of_neighbor_information_from_a_heterogeneous_network_for_discovering_new_drug-target_interactions
3. Baricitinib as potential treatment for 2019-nCoV acute respiratory disease:
[https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(20\)30304-4/fulltext](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(20)30304-4/fulltext)
4. Predicting commercially available antiviral drugs that may act on the novel coronavirus (2019-nCoV), Wuhan, China through a drug-target interaction deep learning model: www.biorxiv.org/content/10.1101/2020.01.31.929547v1.full
5. AI for the repurposing of approved or investigational drugs against COVID-19:
researchgate.net/publication/339998830_AI_for_the_repurposing_of_approved_or_investigational_drugs_against_COVID-19?fbclid=IwAR0UPrNSiYmPSuCkgQiSDI30GufWwKVPa0qhbJG0fPQKY4Zi95z2YIzBxBg

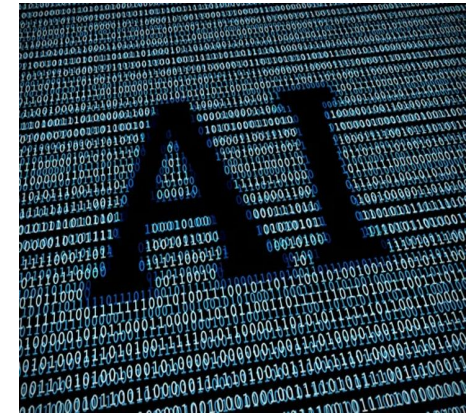
Risk & Prediction:

Medical Home Network Uses AI to Identify High-Risk COVID-19 Patients (Chicago, IL)

- Medical Home Network (MHN) uses existing system of Accountable Care Organizations to target Medicaid patients considered high risk of contracting COVID-19.
- People who receive Medicaid are far more likely to experience barriers of access to health care like social isolation, homelessness, or lack of transportation
- Medicaid recipients are more likely to have a concurrence of multiple health risks contributing to their risk of infection

The MHN Structure

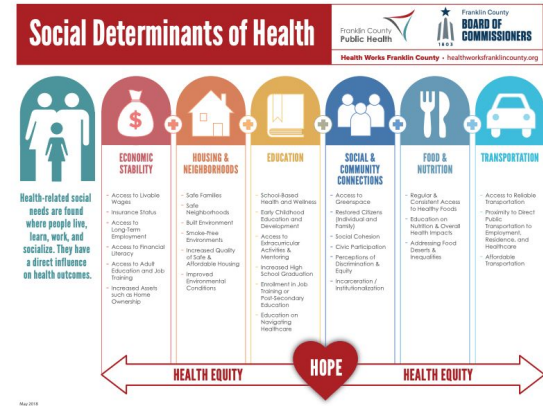
- 120,000 Medicaid pts in Cook County
- 10 Federally Qualified Health Centers
- 3 Hospital systems
- Physician practices



Risk & Prediction:

Medical Home Network Uses AI to Identify High-Risk COVID-19 Patients (Chicago, IL)

- The ACO is working with [ClosedLoop.ai](#), which developed a healthcare data science platform, to use the company's COVID-19 Vulnerability Index—an AI-based predictive model.
- There are 4 categories of data they are using to construct this model:
 - Patient demographics
 - Social Determinants of Health
 - Insurance claims
 - Clinical activity
- Identified patients get outreach from their PCPs - with whom they have an established and trusting relationship - to provide health education.
- ***Not used to predict who will get covid**



ClosedLoop.ai | COVID-19 | SOLUTIONS | PRODUCTS | RESOURCES | COMPANY | GET A DEMO

C-19 index

Open Source Data Science to Fight COVID-19

Prioritizing resources for those most vulnerable to complications from **COVID-19**.

Vulnerability To Severe Illness from COVID-19

4.3X ELEVATED VULNERABILITY

Your scores represent overall risk factors covered by the CDC as creating an elevated risk and your own applicable personal factors should you contract the COVID-19 virus.

Risk Factors

How C-19 vulnerability scores are risk factors based on CDC guidance and national survey-based data.

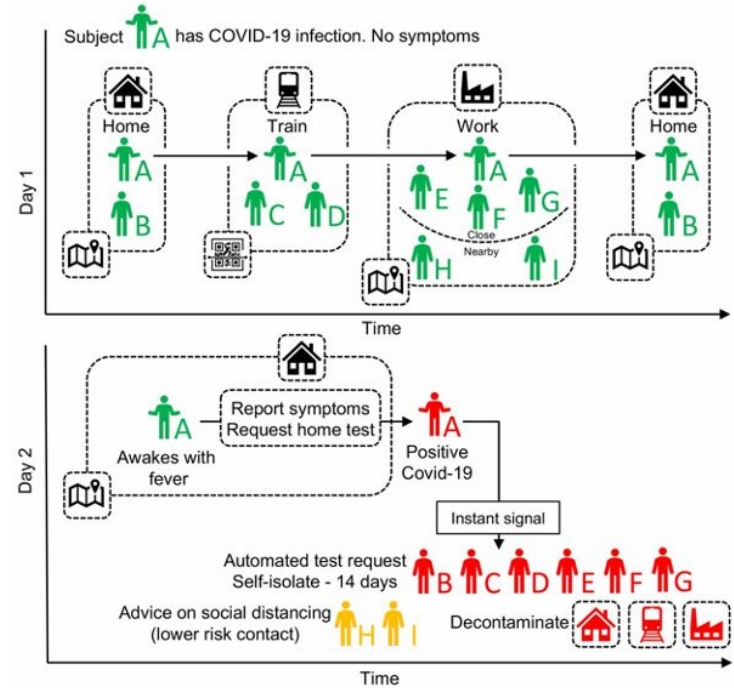
For Organizations
Download the C-19 index on [GitHub](#) or [learn more](#) about how you can use it.

For Individuals
[GET YOUR COVID-19 VULNERABILITY SCORE](#)

Risk & Prediction:

Medical Home Network Uses AI to Identify High-Risk COVID-19 Patients (Chicago, IL)

- App that could be installed on mobile phones in order to alert people who have come into contact with someone who was pre-symptomatic in order to allow them to self-isolate.
- In order to run this technology at population-scale, careful architecture design would be required in order to ensure the **Consistency, Integrity and Availability** of the solution, as well ensuring that the data is not susceptible to unauthorized access.
- **Ingest architecture must be able to scale to well over 6 billion events per day**

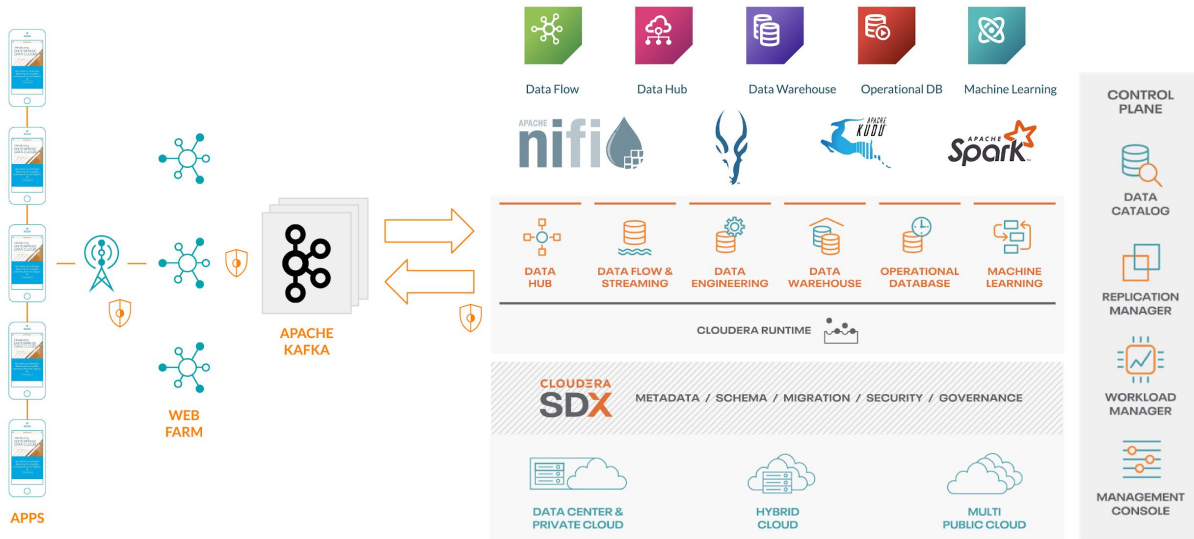


Reference architecture for a contact pairing database that maintains privacy, yet is built to scale to support large-scale lifting of restrictions of movement.

Risk & Prediction:

Medical Home Network Uses AI to Identify High-Risk COVID-19 Patients (Chicago, IL)

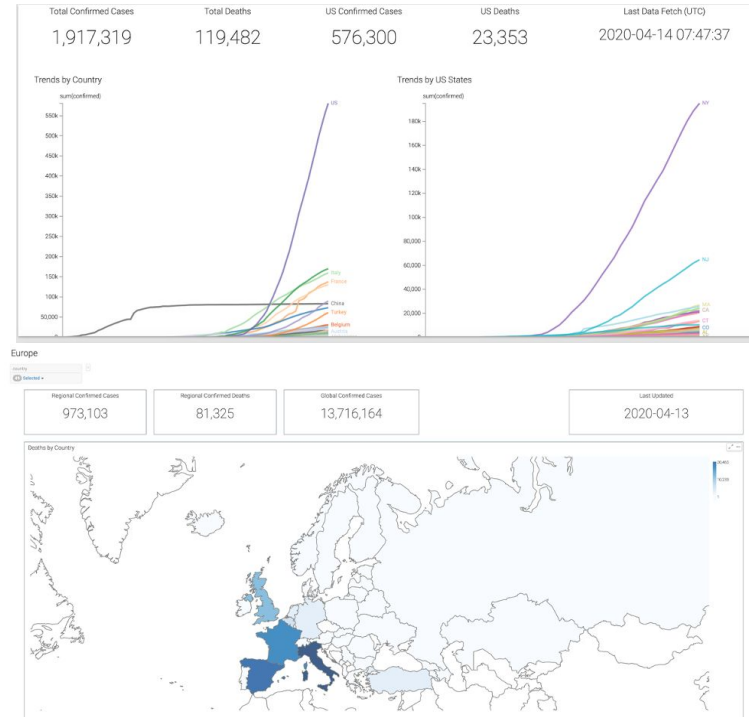
End-to-end Architecture



Risk & Prediction:

Medical Home Network Uses AI to Identify High-Risk COVID-19 Patients (Chicago, IL)

- Machine learning application: predicting customer churn
- Algorithms are typically trained to recognise signals for propensity to churn which usually relate to poor service provision as a result of signal coverage or a superior competing service.
- Case Tracing (aka contact tracing) could be supplemented with similar technology that alerts to higher levels of immediate risk and supports the rapid facilitation of practical interventions such hygiene equipment and deep cleans.
- Once experiments and models are built using CML, CML's Applications feature could be used to provide access to interactive visualizations.



Additional Risk Assessment Projects

Forecasting

1. Carnegie Mellon University, algorithms trained to predict the seasonal flu, are now be re-trained on new data from COVID-19: www.technologyreview.com/2020/03/13/905313/cdc-cmu-forecasts-coronavirus-spread/
2. Institute for the Future of Humanity at Oxford University provides forecasts of the spread of the virus based on the GLEAMviz epidemiological model: www.gleamviz.org/model/
3. COVID - 19 Outbreak Prediction using Machine Learning: www.youtube.com/watch?v=_Hi6_JQesSQ
4. Machine Learning Tools Predict Impact of Quarantine on COVID-19: news.mit.edu/2020/new-model-quantifies-impact-quarantine-measures-covid-19-spread-0416
5. Building a COVID-19 Vulnerability Index: <https://arxiv.org/pdf/2003.07347.pdf>

Tracking

1. Metabiota, a San Francisco-based company, offers an Epidemic Tracker and a near-term forecasting model of disease spread: metabiota.com/sites/default/files/inline-files/Metabiota_Risk_Report_No.3-25Feb2020-COVID-2019_0.pdf
2. Ouraring and UCSF Tempredict study with civilians and front-line workers: <https://ouraring.com/ucsf-tempredict-study>

Looking Forward - ML Adjacent

Ethics and surveillance

1. Covid, AI surveillance, and ethics: www.youtube.com/watch?v=Y5eqenm818w
2. “Even when infections from coronavirus are down to zero, some data-hungry governments could argue they needed to keep the biometric surveillance systems in place”: www.ft.com/content/19d90308-6858-11ea-a3c9-1fe6fedcca75
3. On the responsible use of digital data to tackle the COVID-19 pandemic: www.nature.com/articles/s41591-020-0832-5
4. Covid-19 and the workforce: Critical workers, productivity, and the future of AI: insights.techreview.com/covid-19-and-the-workforce-critical-workers-productivity-and-the-future-of-ai
5. Predict the probability that a patient will pass away within 180 days following his/her last exam: www.synapse.org/#!Synapse:syn18405991/wiki/595482

Calls for proposals and mutual aid

1. Immediate Call for Proposals: AI Techniques to Mitigate Pandemic: c3.ai/c3-ai-microsoft-and-leading-universities-launch-c3-ai-digital-transformation-institute/
2. CovidSafe Contract tracing: covidsafe.cs.washington.edu/
3. Coronavirus Tech Handbook (mutual aid): <https://coronavirustechhandbook.com/home>

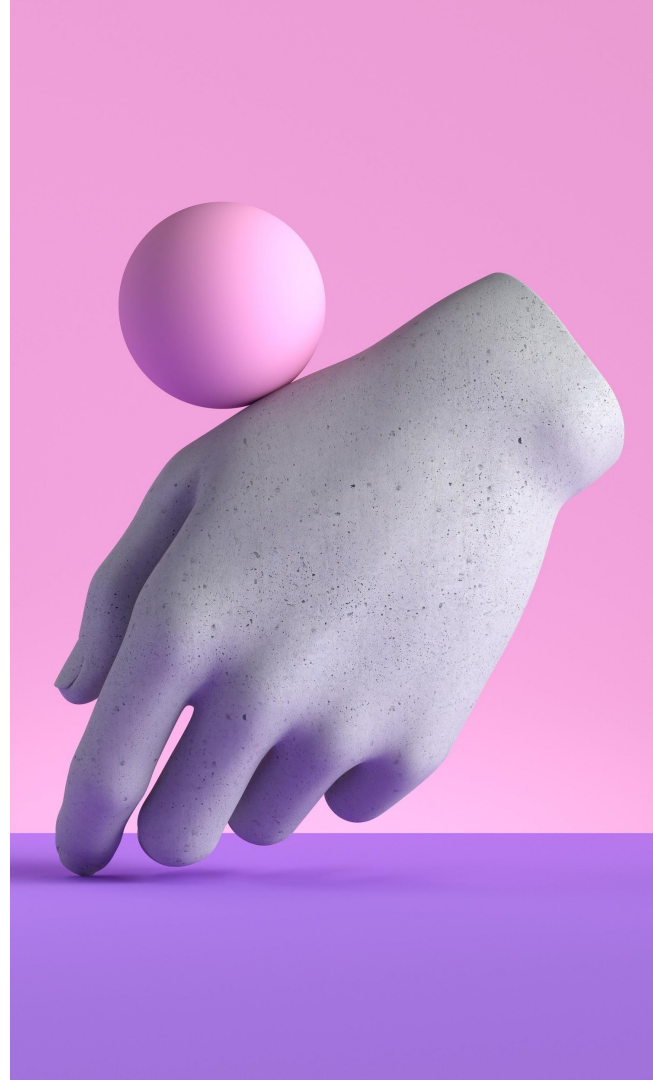
Ambitious ML Applications

- AutoRegressive Integrated Moving Average
- Deep Convolutional Networks
- Generative Adversarial Networks
- Gradient-boosted-tree models
- Deep Reinforcement Learning



Problematic areas for Covid19 ML projects

- Statistical bias in reporting
- Discriminatory data collection
- Incomplete data reporting
- Variance of data
- Community Participation/Trust
- Inaccessibility of data
 - Public understanding
 - Access to proprietary data



Resources:

Dashboards and starter projects



Tableau Covid Data Dashboard



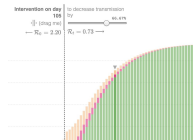
The Covid Tracking Project



Humanitarian Data Exchange



Johns Hopkins Dashboard



Epidemic Calculator (github repo)

Fold@home to donate computer power: <https://foldingathome.org/covid19/>

List of projects on **r/COVIDprojects**: <https://covid19projects.now.sh/>

Python script for analyzing NY Times Covid Data: <https://github.com/tirthajyoti/Covid-19-analysis>

Track Covid cases **with Python**: <https://towardsdatascience.com/how-to-track-coronavirus-with-python-a5320b778c8e>

Resources:

Challenges, datasets

Call to action from the White House: Dataset released 29k articles (13k with full text) between AI2, CZI, MSR, Georgetown, NIH: www.semanticscholar.org/cord19; Scientific questions and text: kaggle.com/allen-institute-for-ai/CORD-19-research-challenge

Datasets

1. The Covid tracking project: <https://covidtracking.com/data>
2. Google Scholar Covid-19 public datasets: console.cloud.google.com/marketplace/browse?filter=solution-type:dataset&filter=category:covid19&pli=1
3. Global research on coronavirus disease (COVID-19) from WHO: www.who.int/emergencies/diseases/novel-coronavirus-2019/global-research-on-novel-coronavirus-2019-ncov
4. Elsevier's Novel Coronavirus Information Center: www.elsevier.com/connect/coronavirus-information-center
5. AWS Covid19 Data Lake: dj2taa9i652rf.cloudfront.net/
6. UW Research on Covid-19 using ML: www.cs.washington.edu/research/covid-19
7. Top 50 R resources on COVID-19: towardsdatascience.com/top-5-r-resources-on-covid-19-coronavirus-1d4c8df6d85f