

Open in app ↗

Sign up

Sign In



Search Medium



Published in Towards Data Science



Wim Naudé

Follow

Apr 1, 2020 · 19 min read · Listen



Save



# Artificial Intelligence against COVID-19: An Early Review

AI has not yet made an impact, but data scientists have taken up the challenge



Source: Pixabay

## Introduction



151



5

COVID-19 disease, caused by the SARS-CoV-2 virus, was identified in December 2019 in China and declared a global pandemic by the WHO on 11 March 2020. Artificial Intelligence (AI) is a potentially powerful tool in the fight against the COVID-19 pandemic. AI can, for present purposes, be defined as Machine Learning (ML), Natural Language Processing (NLP), and Computer Vision applications to teach computers to use big data-based models for pattern recognition, explanation, and prediction. These functions can be useful to recognize (diagnose), predict, and explain (treat) COVID-19 infections, and help manage socio-economic impacts. Since the outbreak of the pandemic, there has been a scramble to use and explore AI, and other data analytic tools, for these purposes.

In this article, I provide an early review, discussing the actual and potential contribution of AI to the fight against COVID-19, as well as the current constraints on these contributions. It aims to draw quick take-aways from a fast expanding discussion and growing body of work, in order to serve as an input for rapid responses in research, policy and medical analysis. The cost of the pandemic in terms of lives and economic damage will be terrible; at the time of writing, great uncertainty surrounded estimates of just how terrible, and of how successful both non-pharmaceutical and pharmaceutical responses can be. Improving AI, one of the most promising data analytic tools to have been developed over the past decade or so, so as to help reduce these uncertainties, is a worthwhile pursuit. Encouragingly, data scientists have taken up the challenge (which implies that the shelf-life of this paper is likely to be brief).

The key take-aways are as follows. I find that AI has not yet been impactful against COVID-19. Its use of AI is hampered by a lack of data, and by too much noisy and outlier data. Overcoming these constraints will require a careful balance between data privacy and public health concerns, and more rigorous human-AI interaction. It is unlikely that these will be addressed in time to be of much help during the present pandemic. Instead, AI may “help with the next pandemic”. In the meantime, gathering diagnostic data on who is infectious will be essential to save lives and limiting the economic havoc due to containment.

## **Actual and Potential Contributions of AI against COVID-19**

There are six areas where AI can contribute to the fight against COVID-19: i) early warnings and alerts, ii) tracking and prediction, iii) data dashboards, iv) diagnosis and

prognosis, v) treatments, and cures, and vi) social control.

### *Early warnings and Alerts*

The case of the Canadian-based AI model, BlueDot, has already become legendary. It illustrates that a relatively low-cost AI tool (*BlueDot* was funded by a startup investment of around US\$ 9 million) can out-predict humans in spotting infectious disease outbreaks. According to accounts, *BlueDot* predicted the outbreak of the infection at the end of 2019, issuing a warning to its clients on 31 December 2019, before the World Health Organization did on 9 January 2020. Researchers working with *BlueDot* also published a notice in the Journal of Travel Medicine on 14 January 2020, where it listed the top 20 destination cities where passengers from Wuhan would arrive. It warned that these cities could be at the forefront of the global spread of the disease.

While *BlueDot* is undoubtedly a powerful tool, much of the publicity it has received contain some exaggeration and some undervaluation of the role of human scientists. First, while *BlueDot* sounded an alarm on 31 December 2019, another AI-based model, HealthMap, at Boston Children's Hospital in the USA, sounded an alarm even earlier, on 30 December 2019. Moreover, only 30 minutes after this, a scientist at the Program for Monitoring Emerging Diseases issued an alert. While the AI-based model was faster by only 30 minutes, it, however, attached a very low level of significance to the outbreak. In essence, it required human interpretation and providing context to recognize the threat. Moreover, even in the case of *BlueDot*, humans remain central in evaluating its output, as Kamran Khan, Founder of *BlueDot*, explains In this podcast. It is therefore rightly stressed that human input, and from various disciplines, is needed for the optimal application of AI.

### *Tracking and Prediction*

AI can be used to track (including nowcasting) and to predict how the COVID-19 disease will spread over time and over space. For instance, following a previous pandemic, that of the 2015 Zika-virus, a dynamic neural network was developed to predict its spread. Models such as these will, however, need to be re-trained using data from the COVID-19 pandemic. This seems to be happening now. At Carnegie Mellon

University, algorithms trained to predict the seasonal flu, are now be re-trained on new data from COVID-19.

Various problems bedevil the accurate forecasting of how the pandemic will spread. These include a lack of historical and unbiased data on which to train the AI; panic behavior which leads to “noise” on social media; and the fact that the characteristics of COVID-19 infections differ from those of previous pandemics. It is not only the lack of historical data but also the problems with using “big data,” e.g., harvested from social media, that have shown to be problematic.

Here, the pitfalls of big data and AI in the context of infectious diseases, as was illustrated in the infamous failure of Google Flu Trends, remain valid. David Lazer, Ryan Kennedy, and Alessandro Vespignani in a 2014 paper in Science referred to these as “*big data hubris and algorithm dynamics*.” For instance, as the infection continues to spread and the social media traffic around it accumulates, so the amount of noise accumulates, which has to be filtered through before meaningful trends can be discerned. For prediction tools that rely on past behaviour, a global outlier event with its mass of new and unprecedented data such as COVID-19 has been described by Ian Rowan as “*the kryptonite of modern Artificial Intelligence approach*” which will affect not only the prediction of infectious diseases, but all prediction models, including those in finance, economics. As he explains, “*many industries are going to be pulling the humans back into the forecasting chair that had been taken from them by the models*”.

One way to deal with big data hubris and algorithm dynamics is through content moderation on social media. The large social media platforms such as Google (YouTube) and Facebook have started to use AI more intensively to do content moderation, including checking for fake news, due to the reduction in human staff resulting from lockdown measures. Relying more on AI for content moderation has laid bare the fact that AI is still doing a poor job of it. YouTube admitted that using AI more extensively in content moderation is “error-prone.” This again illustrates the need for human input to, and direction of, AI.

As a result of a lack of data, too much outlier data and noisy social media, big data hubris, and algorithmic dynamics, AI forecasts of the spread of COVID-19 are not yet very accurate or reliable. Hence, so far, most models used for tracking and forecasting do **not** use AI methods. Instead, most forecasters prefer established epidemiological

models, so-called SIR models, the abbreviation standing for the population of an area that is *Susceptible*, *Infected*, and *Removed*.

For example, the Institute for the Future of Humanity at Oxford University provides forecasts of the spread of the virus based on the [GLEAMviz](#) epidemiological model. [Metabiota](#), a San Francisco-based company, offers an [Epidemic Tracker](#) and a near-term forecasting model of disease spread, which they use to make [predictions](#). Tom Crawford, an Oxford University mathematician, provides a short and concise explanation of these SIR-models in a recent YouTube [video](#).

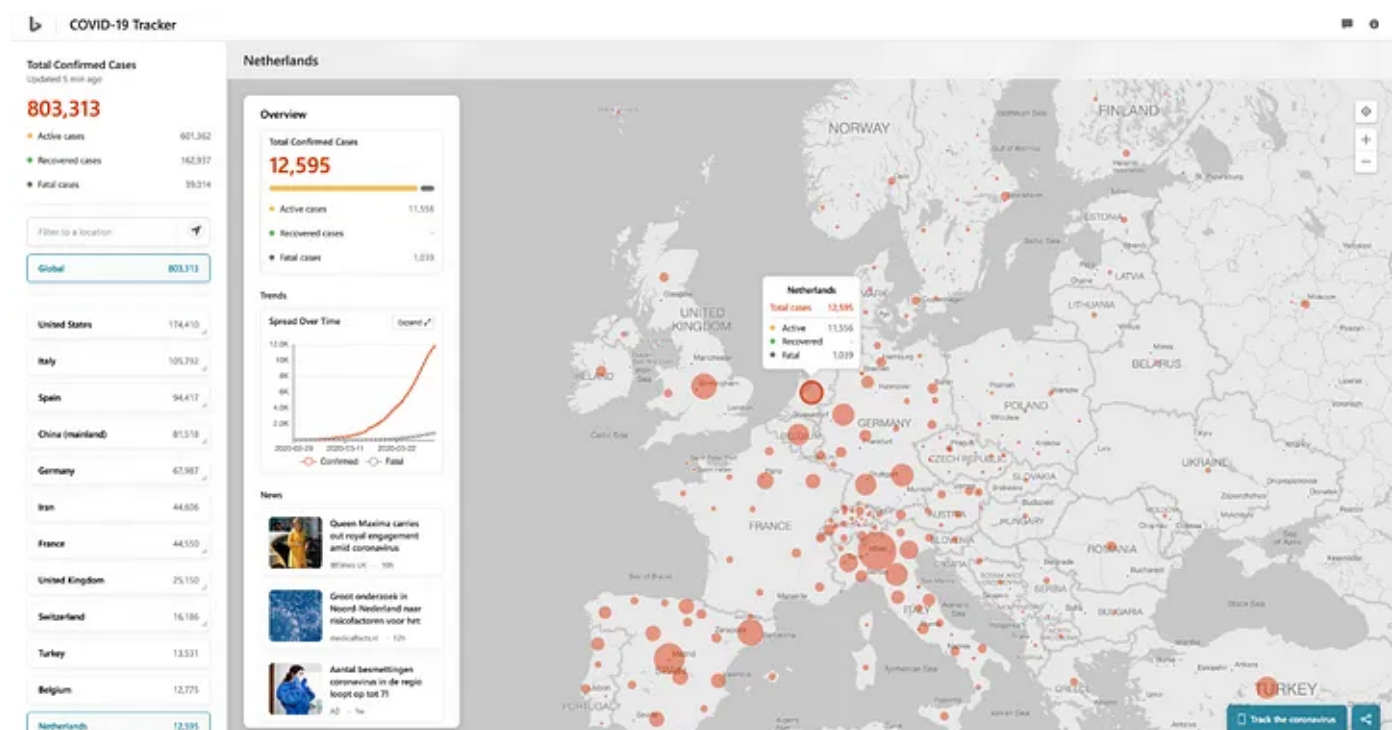
The Robert Koch Institute in Berlin uses an epidemiological SIR model that takes into account containment measures by governments, such as lockdowns, quarantines, and social distancing prescriptions. Their model is explained [here](#) and [here](#). A similarly [extended SIR model](#), taking into account public health measures against the pandemic and using data from China, has recently been pre-published and made available in *R* format. The Robert Kock Institute's model has been used earlier in the case of China to illustrate that containment can be [successful](#) in reducing the spread to slower than exponential rates.

Tracking and predicting the spread of COVID-19 are valuable data inputs for public health authorities to plan, prepare, and manage the pandemic. And to evaluate where they are on the [epidemiological curve](#) and whether they succeed in flattening it. It can also provide rough reflections on the possible impact of measures taken to reduce or slow down the spread. For example, the Robert Koch Institute forecasted that the number of infections in the Netherlands will reach 10,922 by 28 March 2020. At this date, according to John's Hopkins University's [CSSE](#), the total number of infected patients in the Netherlands was lower than predicted, at 8,647. This may strengthen arguments that the government's approach is helping to reduce the growth in infections.

### *Data Dashboards*

The tracking and forecasting of COVID-19 has caused the emergence of an industry of data dashboards that visualizes the pandemic. MIT Technology Review has produced a ranking of these tracking and forecasting [dashboards](#). They rank the top dashboards to be those of [UpCode](#), [NextStrain](#), the John's Hopkins' [JHU CSSE](#), [Thebaselab](#), the [BBC](#),

the [New York Times](#), and [HealthMap](#). Other notable dashboards include [Microsoft Bing's AI tracker](#).



Screenshot of Bing's COVID-19 Tracker, 31 March 2020

While these dashboards give a global overview, an increasing number of countries already have their own dashboards in place; for instance, South Africa established the [COVID 19 ZA South Africa Dashboard](#), which is maintained by the Data Science for Social Impact Research Group at the University of Pretoria.

To facilitate the production of data visualizations and dashboards of the pandemic, Tableau has created a [COVID-19 Data Hub](#) with a COVID-19 Starter Workbook. And [Tirthajyoti Sarkar](#) has published a Python script to illustrate how one could extract data from the New York Times's COVID-19 dataset to create data visualizations of the progression of the infection. [Amanda Makulec](#) calls for responsible visualization of COVID-19 data, listing “Ten Considerations when Visualizing COVID-19 Data”.

### *Diagnosis and Prognosis*

Fast and accurate diagnosis of COVID-19 can save lives, limit the spread of the disease, and generate data on which to train AI models. AI may provide useful input in this regard, in particular with an image-based medical diagnosis. According to a [recent](#)



review of AI applications against COVID-19 by researchers working with UN Global Pulse, studies have shown that AI can be as accurate as humans, can save radiologists' time, and perform a diagnosis faster and cheaper than with standard tests for COVID-19. Both X-rays and Computed Tomography (CT) scans can be used. Adrian Rosebrock offers a tutorial on how to use Deep Learning to diagnose COVID-19 using X-ray images. He makes the point that COVID-19 tests are "*in short supply and expensive, but that all hospitals have X-ray (or CT) machines*". Maghdid et al. (2020) has proposed a technique to use mobile phones to scan CT images.

Several initiatives in this respect are underway. An AI called COVID-Net has been developed to diagnose COVID-19 in chest x-rays using data from patients with various lung conditions, including COVID-19. In China, researchers affiliated to the Renmin University of Wuhan published an AI model (not yet peer-reviewed however) to diagnose COVID-19 from CT scans, concluding that "*The deep learning model showed comparable performance with an expert radiologist, and greatly improve the efficiency of radiologists in clinical practice. It holds great potential to relieve the pressure off frontline radiologists, improve early diagnosis, isolation, and treatment, and thus contribute to the control of the epidemic.*"

Another example of ongoing efforts include that of researchers at the Dutch University of Delft who released an AI model for diagnosing COVID-19 from X-rays. This model, labeled CAD4COVID, is "*an artificial intelligence software that triages COVID-19 suspects on chest X-rays images*". It relies on previous AI models developed by the university of diagnosis of tuberculosis.

The potential of AI in diagnostics is not yet carried over into practice, although it has been reported that a number of Chinese hospitals have deployed "AI-assisted" radiology technologies. Radiologists, however, have expressed their concern that there is not enough data available to train AI models, that most of the available COVID-19 images come from Chinese hospitals and may suffer from selection bias, and that using CT-scans and X-rays may contaminate equipment and spread the disease further. Indeed, the use of CT scans in European hospitals has dropped after the pandemic broke, perhaps reflecting this concern.

Finally, once the disease is diagnosed in a person, the question is whether and how intensively that person will be affected. Not all people diagnosed with COVID-19 will

need intensive care. Being able to forecast who will be affected more severely can help in targeting assistance and planning medical resource allocation and utilization. Researchers at China's Huazhong University of Science and Technology have used ML to develop a prognostic prediction algorithm to predict the likelihood of someone surviving the infection. And a team of researchers from Wenzhou and New York prepared an AI that can predict with 80 percent accuracy which person affected with COVID-19 will develop acute respiratory distress syndrome (ARDS). The sample that they used to train their AI system is, however, small (only 53 patients) and restricted to two Chinese hospitals.

In conclusion, the application of AI to diagnose COVID-19, and to make a prognosis of how patients may progress, has spurred much research effort but is not yet widely operational. As Devan Coldeway concludes, "*No one this spring is going to be given a coronavirus diagnosis by an AI doctor.*" It also seems that comparatively much less effort is on using AI for very early diagnostic purposes, for instance, in identifying whether someone is infected before it shows up in X-rays or CT scans, or on finding data-driven diagnostics that have less contamination risk.

### *Treatments and Cures*

Even long before the COVID-19 outbreak, AI was lauded for its potential to contribute to new drug discovery. In the case of COVID-19, a number of research labs and data centers have indicated that they are recruiting AI to search for treatments for and a vaccine against COVID-19. The hope is that AI can accelerate both the processes of discovering new drugs as well as for repurposing existing drugs.

For example, Google's DeepMind has predicted the structure of the proteins of the virus — information that could be useful in developing new drugs. However, as DeepMind makes clear on its website, "*we emphasize that these structure predictions have not been experimentally verified...we can't be certain of the accuracy of the structures we are providing.*"

Researchers from South Korean and the USA have published results from using ML to identify an existing drug, *atazanavir*, which could potentially be repurposed to treat COVID-19. Researchers at Benevolent AI, a UK AI startup, and Imperial College have published a paper in The Lancet, identifying *Baricitinib*, a drug used for rheumatoid



arthritis and myelofibrosis, as a potential treatment for COVID-19. Researchers attached to Singaporean firm *Gero*, using a deep neural network, identified a number existing experimental and approved drugs, including *Afatinib*, a lung-cancer treatment, that could potential be used to treat COVID-19. –Their paper, however, has not yet been peer-reviewed.

It is not very likely that these treatments (and perhaps cures) will be available in the near future, at least to be of much use during the current pandemic. The reason is that the medical and scientific checks, trials, and controls that need to be performed before these drugs will be approved, once they have been identified and screened, will take time- according to estimates up to 18 months for a vaccine.

### *Social Control*

AI has been, and can further be used, to manage the pandemic by scanning public spaces for people potentially infected, and by enforcing social distancing and lockdown measures. For example, as described in the South China Morning Post, “*At airports and train stations across China, infrared cameras are used to scan crowds for high temperatures. They are sometimes used with a facial recognition system, which can pinpoint the individual with a high temperature and whether he or she is wearing a surgical mask*”.

Chinese firm Baidu is one of the producers of such infrared cameras that uses computer vision to scan crowds. It is reported that these cameras can scan 200 persons per minute and will recognize those whose body temperature exceeds 37,3 degrees. Thermal imaging has however been criticized as being inadequate to identify from a distance a fever in people who are wearing glasses (because scanning the inner tear duct gives the most reliable indication) and because it cannot identify whether a person’s temperature is raised because of COVID-19, or some other reason.

More worryingly, as the South China Morning Post further reports, “*This system is also being used to ensure citizens obey self-quarantine orders. According to reports, individuals who flouted the order and left home would get a call from the authorities, presumably after being tracked by the facial recognition system.*”

This usage is not limited to China. An AI-based computer vision camera system scanning public areas has been used to monitor whether people in the UK city of Oxford keep to the social distancing measures of the government. A USA computer vision-based startup is already offering “social distancing detection” software, which uses camera images to detect when social distancing norms are breached, after which it will send out a warning. In more extreme cases, the Israeli government has approved cyber-monitoring by its security services to identify and quarantine people that may be infected, and Russia is rolling out a combination of an app and QR system to track infected people and control movement.

Whereas using AI to predict and diagnose COVID-19 is hampered due to lack of historical training data, AI tools such as computer vision and robots are not. Therefore, we are more likely over the short term to see this type of AI being used and used moreover for social control. Related technologies, such as mobile phones with AI-powered apps or wearables that harvest location, usage, and health data of their owners, are also more likely to be employed. Georgios Petropoulos at Bruegel states that such apps “*enable patients to receive real-time waiting-time information from their medical providers, to provide people with advice and updates about their medical condition without them having to visit a hospital in person, and to notify individuals of potential infection hotspots in real-time so those areas can be avoided.*”

Based on data from mobile devices, Google has made available “COVID-19 Community Mobility Reports”, available for 131 countries, which allows one to observe the impact of containment measures on people’s mobility.

Screen shot from Google's COVID-19 Community Mobility Report for the Netherlands

Useful as these are, the fear is that once the outbreak is over, that erosion of data privacy would not be rolled back and that governments would continue to use their improved ability to survey their populations- and use the data obtained in the fight against COVID-19 for other purposes. Yuval Noah Harari (2020) warns *“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 because they fear a second wave of coronavirus, or because there is a new Ebola strain evolving in central Africa, or because . . . you get the idea.”*

### **Constraints: Too Much, and Too Little, Data**

AI has the potential to be a tool in the fight against COVID-19 and similar pandemics. However, as Georgios Petropoulos at Bruegel concludes, *“AI systems are still at a preliminary stage, and it will take time before the results of such AI measures are visible.”* It has been shown here that the current use of AI is actually constrained by, on the one hand, by a lack of data, and on the other hand, by too much data. There is a lack of historical data on which to train AI models, not enough open datasets and models to

work on, but also the potential problems of big data hubris, non-adjustment of algorithms, and a deluge of scientific findings and outlier data which need to be shifted and evaluated before eventually being put through clinical trials.

In contrast, where AI is easier to use, such as in surveillance, we are likely to see more effort — but with potential adverse longer-term consequences for privacy and related human rights concerns. In what follows, I will deal in more detail with these matters.

First, as far as the need for more data is concerned, more new training data is explicitly needed on COVID-19; more openness and sharing of information is required, and more collaborative and multidisciplinary research is necessary to improve the ability of AI. Moreover, more diagnostic testing needs to be done. In all of these, the role of humans in interacting with and steering AI is necessary.

So far, there has been promising progress with a number of notable activities recognizing the importance of building and sharing existing datasets and information about the epidemic. One of the first has been the World Health Organization's (WHO) Global Research on Coronavirus disease database, with links to other similar initiatives.

One of the most ambitious of these focusing on AI, is perhaps the joint initiative between Semantic Scholar, the Allen Institute for Artificial Intelligence, Microsoft, Facebook, and others, to make openly available the COVID-19 Open Research Dataset (CORD-19) which contains around 44,000 scholarly articles which are now available for data mining.

Relatedly Kaggle, a data science competition platform, has issued a data competition based on this data, a “COVID-19 Open Research Dataset Challenge”. Zindi, Africa's largest data competition platform, has similarly launched a competition to “accurately predict the spread of COVID-19 around the world over the next few months”.

*Elsevier* has made publicly available in its Novel Coronavirus Information Center early-stage and peer-reviewed research on COVID-19 and to around 20,000 related articles on ScienceDirect, as well as the full texts for data mining. Similarly *The Lens* has made available all its data on patents in what it calls the Human Coronavirus Innovation Landscape Patent and Research Works Open Datasets to support the search for new and repurposed drugs. *Google* has made available (until 15 September 2020) COVID-19

Public Datasets on its Cloud Platform, and *Amazon* has launched a public [AWS COVID-19 data lake](#), which it describes as “a centralized repository of up-to-date and curated datasets on or related to the spread and characteristics of the novel corona virus (SARS-CoV-2) and its associated illness, COVID-19”.

Other data-gathering and open innovation initiatives include that of The University of California, Berkeley, the University of Illinois at Urbana-Champaign, and C3.ai who established the [C3.ai Digital Transformation Institute](#). This Institute has launched a Call for Proposals for “*AI Techniques to Mitigate Pandemic*.” These should deal amongst others with “Applying machine learning and other AI methods to mitigate the spread of the COVID-19 pandemic”, and “Data analytics for COVID-19 research harnessing private and sensitive data”. Open access data is also gathered and made available by the [GISAID Initiative](#) (formerly the Global Initiative on Sharing All Influenza Data).

It is not only the large tech companies, publishers, and universities that are promoting open access to data and scientific literature on COVID-19, but also smaller startups and NGOs. For example, [Newspeak House](#) — a UK based independent residential college — has started a crowdsourcing initiative, a [Coronavirus Tech Handbook](#), to which it has invited the public to contribute. And [Emily Chen](#) and colleagues published the first public COVID-19 Twitter dataset.

It is not only a lack of data that constrains AI applications, but also, perhaps paradoxically, too much data. As was noted, as the pandemic progresses and the issue dominates the news and social media, too much big data noise and outlier data is created, and algorithms will be overwhelmed — this was the lesson from the *Google Flu Trends*’ failed initiative. Content curation and algorithmic adjustment, both involving human common sense, become especially valuable then. Furthermore, scientists will need to deal with the deluge of scientific papers and new data being generated and shift through these.

More than [100 scientific](#) articles on the pandemic now appear daily. This potential information overload is, however, where data analytic tools can play an important role. An example of an initiative in this regard is the [COVID-19 Evidence Navigator](#), which provides computer-generated evidence maps of scientific publications on the pandemic, daily updated from [PubMed](#).

Screenshot of Gruenwald et al.'s COVID-19 Evidence Navigator, 1 April 2020

## Conclusions

AI is not yet playing a significant role in the fight against COVID-19, at least from the epidemiological, diagnostic and pharmaceutical points of view. Its use is constrained by a lack of data and by too much noisy and outlier data. The creation of unbiased time series data for AI training is necessary. A growing number of international initiatives in this regard is encouraging; however, there is an imperative for more diagnostic testing. Not only for providing training data to get AI models operational, but moreover for more effectively managing the pandemic and reducing its cost in terms of human lives and economic damage.

At the time of writing, the significant efforts of all affected countries have been to shut down their economies through lockdowns, enforcing social distancing, and canceling events. These measures seem, for now, to have succeeded in slowing down the spread. However, whether these measures are sustainable for more than a couple of weeks is doubtful. According to the Imperial College COVID-19 Response Team, *“The major challenge of suppression is that this type of intensive intervention ... will need to be maintained until a vaccine becomes available, given that we predict that transmission will quickly rebound if interventions are relaxed.”*

More diagnostic testing will be helpful to eventually halt the pandemic, limit the economic damage from lockdowns, and avoid a rebound once restriction are relaxed.



Mathias Dewatripont and colleagues make a case for extensive diagnostic testing of the population to allow people to return to work only if they are not infectious, to place in quarantine those who are. They also call for more randomly sampled tests in order to improve our estimates of the proportion of the population with the virus that remain asymptomatic. At present, we just do not know how many people are infected. In essence, it may be, as a study in Science suggests, that 86 percent of all infections may be undocumented. If this is the case, then the danger of a rebound of the pandemic is highly likely. Thus, overcoming limited data in terms of who is infectious is critical.

Finally, data is central to whether AI will be an effective tool against future epidemics and pandemics. The fear is, as I already mentioned, that public health concerns would trump data privacy concerns. Governments may want to continue the extraordinary surveillance of their citizens long after the pandemic is over. Thus, concerns about the erosion of data privacy are justified.

A full discussion of the legal and ethical dimensions of data management falls outside the scope of this article. Two excellent recent commentaries are, however, published in Bruegel and Nature. In short, given the public health threat posed by the pandemic, the European GDPR (Article 9) allows personal data collection and analysis, as long as it has a clear and specific public health aim. Flexibility to gather and analyze big data promptly is essential in combatting the pandemic, even if it may require that the authorities collect more personal data than many people would feel comfortable with. Therefore, it is crucial that the authorities take particular care in their handling of such data and their justifications and communications to the public at large. The danger is that the people could lose trust in government, which will, as Ienca and Vayena pointed out, “*make people less likely to follow public-health advice or recommendations and more likely to have poorer health outcomes.*”

## Sign up for The Variable

By Towards Data Science

Every Thursday, the Variable delivers the very best of Towards Data Science: from hands-on tutorials and cutting-edge research to original features you don't want to miss. [Take a look.](#)

By signing up, you will create a Medium account if you don't already have one. Review our [Privacy Policy](#) for more information about our privacy practices.



Get this newsletter

[About](#) [Help](#) [Terms](#) [Privacy](#)

Get the Medium app