apheris AI

Data privacy for COVID-19 smartphone apps

31 March 2020

Abstract

The previously-unknown respiratory tract disease COVID-19 is spreading all over the world and overloads or will soon overload health care systems. In order to manage the situation, governments need models to evaluate possible measures such as social distancing. After introducing disease models, we give an overview on measures to fight COVID-19 such as more testing, monitoring and self-isolation. The latter two measures, monitoring and self-isolation, need more and precise data, e.g., about infections in different groups of society and possible contacts of newly diagnosed patients. A smartphone app can help to collect this data and communicate the results faster. However, three key challenges need to be solved: First, data must be collected from reliable sources, second, privacy threats should be taken into consideration and mitigated using state-of-the-art mechanisms such as differential privacy or private set intersection, and third, the implementation needs not only to be fast, but also correct and efficient. We at apheris AI are a team of leading experts at the intersection of data privacy, cryptography and biomedical data science and work with the world's largest pharmaceutical and telecommunication companies on privacy preserving artificial intelligence. We offer our knowledge in suitable privacy tools to fight the COVID-19 pandemic without violating basic privacy rights of individuals.

1 Introduction and medical background

At the beginning of the year, reports of an outbreak in China of a previously-unknown respiratory tract disease with the causative agent being a virus emerged. This event is fundamentally changing life as we know it, in almost all parts of the world: a pandemic with still no foreseeable end.

SARS-CoV-2 is the causative agent of the pandemic outbreak. It is a newly encountered member of the coronavirus family which belongs to the RNA-viruses, is in its behaviour comparable to influenzaviruses or SARS-CoV — the causative agent of the pandemic outbreak 2002/03 [1][2]. As soon as virus particles get into a host (human), they start invading cells (in this case predominantly respiratory tract cells), and the host's cells replicate the virus's genome. Virus particles get into the host's saliva and humans infect each other by talking to infected individuals, by touching hands and by close face-to-face interaction [2][3]. A number that is a good landmark for the transmission rate, or the "infectiousness" of any infectious disease, is the basic reproductive number (R_0): How many cases will be expected to be infected by a single positive individual in a susceptible

population? Our current numbers from China suggest a range between 1.4 and 3.9 for the transmission rate [3][5]. A second important parameter is the case fatality rate, or the "virus-associated mortality". A low fatality rate and many asymptomatic (but infectious) cases increases the basic reproductive number [2] [4] [6].

Symptoms of coronavirus disease (COVID-19) are widespread: from asymptomatic patients to patients with flu-like symptoms up to a severe pneumonia leading to a severe acute respiratory distress symptom (ARDS)[1][2], making the ventilation of patients inevitable. Considering the low fatality rate [1] and its age adaptation, it is tempting to ease personal concerns about the new infectious disease but this can lead to a fatal outcome: Spreaders of the disease are young, healthy individuals with no or light symptoms [5]. But in a decompensating health system, not only the elderly and multimorbid individuals will lose out.

Given the lack of reliable and long-term data regarding incubation period, virulence, contagiousness, and other transmission parameters [1] for the novel coronavirus SARS-CoV-2 and the lack of reliable drugs and vaccines [3], containment measurements [3] remain the only feasible option to face the ongoing outbreak of the virus that is leading to a collapsing health system with thousands of deaths, as seen in hotspots.

2 Modelling the COVID-19 pandemic

In order to understand the impact of social distancing, self isolation and other restrictions, we need to simulate the spread of an epidemic. We first introduce a simple epidemic model called SIR (**S**usceptible, **I**nfected, **R**ecovered). Then we add new features to the model to obtain a better description of a real epidemic. We show the impact of different containment measures to understand how a smartphone app can help to reduce the impact of the virus on the daily life.

2.1 The standard SIR model

The SIR model is one of the simplest models to describe the spread of an epidemic [7]. The model consists of 3 categories: **S** for the susceptible people, **I** for the infected people and **R** for the sum of recovered and deceased people. The classic SIR model is described by 3 ordinary differential equations:

$$\begin{aligned} \frac{dS}{dt} &= -\beta \frac{IS}{N} \\ \frac{dI}{dt} &= \beta \frac{IS}{N} - \gamma I \\ \frac{dR}{dt} &= \gamma I \end{aligned}$$

The condition $S + I + R = N$, where N is constant and equal to the total population, is valid at any time. The three equations written above can be interpreted in the following manner:

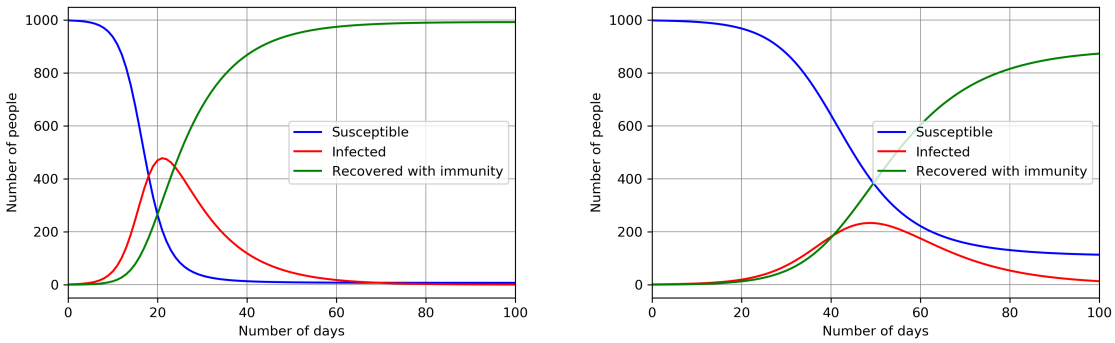

Figure 1: The evolution of an epidemic using the SIR model, with $\gamma = 0.1$ and $\beta = 0.5$ on the left panel and $\beta = 0.25$ on the right panel

- at the beginning of the epidemic the entire population is susceptible to the infection (**S**). If there is a single infected person, other people can get the infection, going from the category **S** to the category **I**. The strength (speed) of the spread of the virus is determined by the parameter β;
- the number of infected people increases when susceptible people get infected. After a typical timescale equal to $1/\gamma$ infected people **I** go in the third category **R**;
- the category **R** includes the sum of people that recovered or died after infection.

The classic SIR model depends only on two parameters, β and γ. Both β and γ have dimension of time^{-1}. From here on we will use the days as units of time. The basic reproductive number R_0 can be obtained by combining the previous parameters and it is dimensionless:

$$R_0 = \frac{\beta S}{N\gamma}$$

R_0 gives immediately information on the spread of the epidemic. If $R_0 \leq 1$ the epidemic will stop spontaneously, while with $R_0 > 1$ it will continue spreading. In Fig. 1 we show two different evolutions of the pandemic to demonstrate the effect of different values of β. Both simulations start with the initial condition of 1 infected person, 1000 people in total, fixing $\gamma = 0.1$ (i.e. typical duration of the illness of 10 days). In the simulation on the left we set $\beta = 0.5$ and on the right we assume $\beta = 0.25$, i.e. when social distancing measures are taken. We notice that in the left panel the peak of infected people comes after 20 days, with roughly half the population infected. On the right panel the peak is shifted and comes after 50 days and the number of infected people at the peak is less than half. This shows the importance of social distancing, since social distancing and quarantines reduce the parameter β, helping in reducing the number of infected people at the peak of the epidemic. This is indispensable to avoid the collapse of the healthcare system and especially intensive care units.

In the classic SIR model the parameter β is constant, therefore it cannot account for the effect produced by a quarantine. Moreover both recovered and deceased go in the same category and this does not permit to forecast the number of deaths at the end of the epidemic. For this reason it is useful to go beyond the classic SIR model, as explained in the next section.

2.2 Beyond the classic SIR model

In the classic SIR model the parameter β is constant in time. This means that it cannot account for a slowdown of the spread due to the quarantine. To simulate a more realistic scenario, it is necessary to go beyond the classic SIR model. We denote this model as SIR 2.0, implemented by A. Palladino as open source code [8]. Compared to the classic SIR it contains the following new features:

- the parameter β changes in time, to account for effects due to quarantine. Particularly $\beta = \beta_0$ below a time t_0, while it varies as:

  $$\beta(t) = \beta_0 \, e^{-(t-t_0)/\beta_1} \quad \text{for } t \geq t_0$$

  The time t_0 represents the starting time of the containment effect. β_1 has the dimension of time;

- the class **R** is now divided in deaths (D) and recovered with immunity ($\tilde{R}$), introducing a dimensionless death rate (d_r):

  $$\begin{aligned} \tilde{R} &= (1 - d_r)R \\ D &= d_r R \end{aligned}$$

- we take into consideration the possible presence of asymptomatic patients. The virologist Ilaria Capua has suggested that $\frac{2}{3}$ of patients in Italy might be asymptomatic [9]. A different study, conducted by the Italian Foundation GIMBE, found a similar result [10].

to the classic SIR there are 3 more parameters, β_1, t_0 and d_r.

2.3 The Italian case

Using the model SIR 2.0 it is possible to describe the spread of COVID-19 in Italy with an average error of 5%. This has been done for the website [11] that is updated on a daily bases. The Italian data, updated to the 28th of March, are well described by the following set of parameters:

$$\beta_0 = 0.415, \ \beta_1 = 28.3, \ \gamma = 1/14, d_r = 6.5, t_0 = 17$$

assuming $t = 1$ as first day of the epidemic. The comparison between the model SIR 2.0 and the real data is reported in Fig. 2.

On the left panel we show a comparison between the total number of infected, recovered and deaths (dashed curves for the model, squares for real data), assuming that $\frac{2}{3}$ of patients are asymptomatic. There is a good agreement between the number of infected people (with symptoms) and the number of deaths. The lack of agreement among recovered with immunity is justified by the fact that recovered asymptomatic patients cannot be included in the real data, but only in the model. As result of the simulation, the peak of the epidemic is expected for mid-April in Italy.

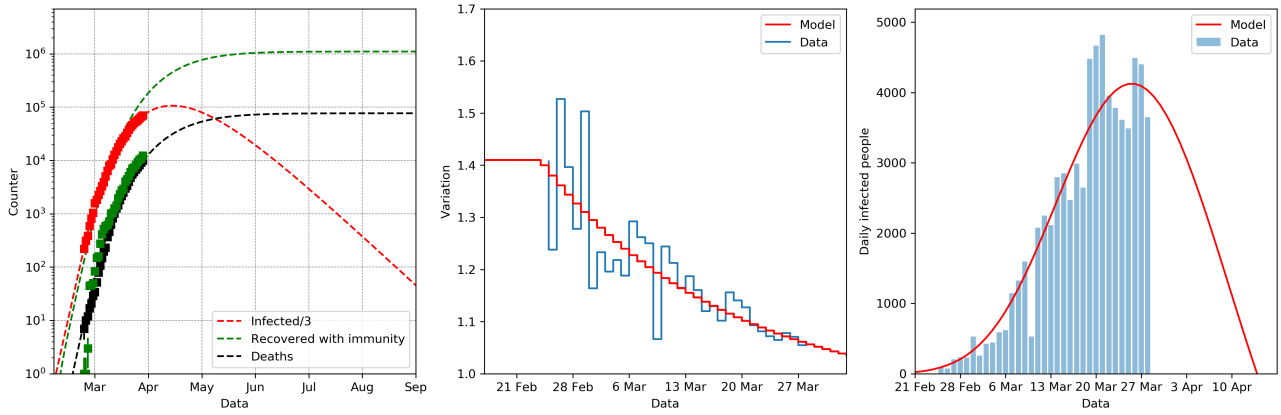

Figure 2: Simulation of the Italian case using the SIR 2.0

The central panel shows the ratio between the total number of cases at day t and day $t-1$. The blue function represents the true data, while the red data represents the model. We notice a good agreement in the behavior, except for statistical fluctuation that cannot be taken into account in the SIR 2.0. The incremental rate started to decrease at the end of February, when restrictive measurements have been taken by the Italian government. This shows the importance of self-isolation.

On the rightmost panel we show the number of new infected people. Also in this case there is a good agreement between the behavior of real data and the behavior of the model. We are right now on the peak of newly-infected people. This means that in the next days the pandemic is expected to slow down. However it does not mean that the number of total infected will decrease, since this will happen when the number of new infected people will be smaller than the number of recovered ones. This is unlikely to occur before 2 weeks in Italy, as explained before.

3 Fighting against the uncontrolled spread of COVID-19

It is vital to prevent the hospitals — and in particular intensive care units — from being overwhelmed by this pandemic. In Italy, as shown in section 2.3, roughly 100,000 people are expected to be infected at the peak of the pandemic. In the current outbreak in Italy, 12 percent of all detected COVID-19 cases were admitted to the intensive care unit [12], which represents a demand for intensive care for 12,000 people. However, according to a 2012 study, Italy had 12.5 intensive care unit beds per 100,000 people [13], i.e. only 7,500 beds. A similar situation might happen in Germany and in other countries. Therefore the imposed restrictions on our society to reduce social contacts are necessary.

In order to minimize the impact of the pandemic to economy and the health care system, the spread of COVID-19 must be measured and controlled. In order to do that, we need to ramp up

COVID-19 testing, we need technology and algorithms that empower us to understand how the virus is spreading and we need to empower every person at risk to self-isolate.

3.1 Ramping up testing programs

As suggested by the WHO in an official statement [14], massive testing with RT-PCR (gold standard diagnostics for most viral agents) in a susceptible population should be conducted as a basis for containment strategies. This is even more important for a disease like COVID-19, which first of all shares symptoms with other virus-driven diseases like the seasonal flu or the common cold and secondly that remain asymptomatic in a percentage of individuals. Symptoms may be misunderstood in individuals, leading to incorrect numbers of infected individuals or to incorrect isolation strategies [15].

3.2 Monitoring and slowing the spread of the disease

A precise understanding of the spread of the virus is indispensable for governments. Political decisions that are data and science-driven is the only chance to come to the right decision and minimize the impact on the population and the economy. Asian countries like Singapore, China or South Korea have shown that tracking and controlling the movement of people via smartphone apps is effective in slowing down the spread of the disease. These smartphone apps closely track users' locations and cross-reference them with lists of patients that have been diagnosed with COVID-19. However, the infringement of these apps on the individual's right is tremendous and we strongly position ourselves against such mass surveillance. We need a system to closely monitor the movement of the population while fully respecting the privacy of each citizen.

3.3 Empowering citizens for high precision self-isolation

In order to minimize the negative impact on the society and economy, we need to develop solutions that empower the people for high-precision self isolation. People at risk need to be able to avoid COVID-19 hotspots and citizens need to be informed and alerted if they have been exposed. At some point, it will become essential to allow citizens to navigate and slowly re-enter public spaces again. We need to develop technical solutions that reduce the epidemic threat and minimize the effect on the economy.

4 Call for COVID-19 smartphone app to fight global pandemic

Individuals need information to balance self-isolation and re-socialization, and government officials need to monitor and understand the disease spread to decide on regulations and manage the

situation. Both citizens and government officials will profit from more data availability and one or several smartphone apps are needed. We list several concrete scenarios below.

4.1 Use cases for a COVID-19 app

Fast contact tracing: Finding contacts of infected COVID-19 patients and isolating them is one of the best measures to slow down the disease spread as we explained above in section 3.2. Currently that is done manually by trained professionals. The process is slow and insufficient as the capacities of professionals are exceeded. Even worse, the process relies on COVID-19 patients' memories of whom they might have met, and they will not remember strangers they met randomly in public places. Having additional location traces collected by the patients' phones might lead to more accurate data and thus more effective contact tracing. Additionally, a COVID-19 app could help notify contacts faster and thus lead to faster quarantine decisions.

Faster and easier communication: Under the assumption that a large part of the population uses the app, information can be distributed via this trusted channel and reach citizens earlier. Be it requests for quarantine due to an infected contact or a fine-grained isolation request, individuals will only follow these requests if they are delivered via a trusted channel. Fine-grained isolation requests could for example notify key workers (e.g. healthcare or logistics personnel) to quarantine in order to be available when current front-line workers get sick, or protect high-risk groups by preventive self-isolation.

More and accurate data on disease spread: Given that health care providers are already overloaded with raising infections, patients with mild symptoms might choose not to seek health care. High-risk citizens on the other hand might choose to self-isolate in order to reduce infection risks and severe outcomes. Knowing how many individuals choose this path and whether they properly self-isolate is key to accurate data on disease spread. An interface that enables patients to track their symptoms is therefore necessary. An app developed and distributed by trusted authorities must be a solution.

Proof of health: Measures like quarantine, rapid testing and contact tracing will lead to lower infection rates such that workers can return to work and public places such as schools and universities can open again. Knowing some workers are healthy and can return to their work place without risk allows an earlier re-start, however, testing capacities are limited. If an individual can prove with their location traces on her smartphone that they properly self-isolated and were not in contact with any infected patient a re-enter into public space can be accelerated and economic outcomes are improved.

4.2 Key steps and challenges

The above mentioned use cases all share three basic steps:

1. Data acquisition: Relevant data such as location traces needs to be collected.
2. Computation of statistics or recommendations based on the data: e.g., has an infected individual met another individual who should therefore self-isolate or how many health care workers are infected.
3. Communication of the results: Finally, the computation result needs to be post-processed into a form that is easy to access for the recipient, e.g., a short message for an isolation request or a detailed statistic for a government official. The result needs to be delivered over a secure channel in order to establish trust in the message itself.

Independent of the use case, we identify three key challenges that need to be solved for any COVID-19 App:

Challenge 1: Reliable, up-to-date data sources that can not be faked and that include large enough parts of the society Tamper-proof data acquisition is necessary since people might have reasons to fake data, be it due to panic or perceived benefits. For example, if a "Proof of Health" would be implemented, individuals are incentivized to spoof their location in order to return to work earlier, however, risk of further infection spread increases dramatically. Furthermore, we should take into consideration that vulnerable groups might not be part of the app users, e.g., elderly people who do not use a smartphone.

Challenge 2: Identification of privacy threats and choice of suitable privacy-preserving technologies This is crucial for trust and wide-spread usage in western societies. We list some threats and solutions in the next sections.

Challenge 3: Correct, reliable implementation of the chosen methods with given hardware, software and time constraints and a large enough user base The best privacy technology is worthless if implemented incorrectly, or if it consumes too many resources to be usable. Furthermore, if not enough data can be collected, valid conclusions cannot be made. One key component for wide-spread user adoption is a privacy guarantee, as we motivated before.

5 Technology for privacy preserving computations on mobile devices

The first step towards a privacy solution is the identification of privacy threats. Once the threats are identified, mitigation techniques that best fit to the problem can be chosen and tested with the threat model in mind.

As explained above, *location traces* are necessary for contact tracing, however, they are a very sensitive data type [16]. Knowing the location traces of an individual leaks the home location to a potential attacker, knowing in addition that the victim is currently not at home, can for example be used for burglaries. Even worse, when the target is not under quarantine, location traces may also leak habits, which can be used for targeted advertising, spear-phishing or even blackmail. The location traces of more than one target additionally contain relationship information, since the attacker can infer who meets whom. Therefore, location traces should be handled with care. Storage at centralized servers can be a valuable target for hacks. Storing the data on the users' devices instead offloads the burden of data protection to the individual, who needs to check that other applications running on the same device do not sniff the data. Encrypted data storage might be a first line of defense against such threats.

Compared to location traces, *fine-grained statistics* seem to be less privacy-sensitive. Nevertheless, the specific use-case needs to be studied to ensure no membership inference attacks are possible [17]. In a nutshell, membership inference attacks only establish the membership status of a target in a group, but the metadata of the group might be a privacy threat. In the COVID-19 case, metadata might be age range, existing medical preconditions or smoking status, and all this information may leak should membership inference attacks be launched successfully [18].

Depending on which *data types* are used as a proxy, new privacy risks arise. For example, practitioners suggest to use purchase histories as a proxy for location to enable contact tracing. However if this data is not preprocessed with care, not only the point of sale is communicated, but also the purchase itself, which can again be used for targeted advertisements, spear-phishing or blackmailing. Further, medical data is crucial for the correct modeling of the disease, but should not leak any pre-existing conditions of the individual.

Privacy research has developed several tools that are helpful for the above mentioned threats. *Cryptographic solutions* are based on mathematically hard problems and ensure privacy by enabling easy computation in the one direction, e.g., from data to output, that are hard to invert, i.e., it is almost impossible to compute back to the data knowing only the output. This comes at a price of higher computation and communication overhead, but yields accurate results.

Differential Privacy on the other hand ensures privacy by adding randomness to the summarized result such that the individual's contribution is "hidden" by the noise [19]. This, of course, decreases accuracy of the results and privacy parameters need to be carefully tuned in order to get a good trade-off between privacy and accuracy. Compared to cryptographic solutions, differentially private algorithms scale more easily to larger amounts of data.

Depending on the use case, the right tool needs to be used. When high accuracy is necessary, for example for location traces, the noise added by a differentially private algorithms might be unacceptable. Cryptographic tools for private set intersection may thus be more suitable for contact tracing [20]. Statistics about infection rates and quarantine compliance however might be a case where differential privacy is the right tool of choice. Under the assumption that the statistics were computed from a reasonably sized data set, the noise added becomes negligible and does not disturb the interpretation of the data, e.g., whether infection rates grow or shrink.

Finally, we need to point out that privacy alone is not sufficient. We also need trustworthy data sources to ensure the results and conclusions are based on correct data. Notice that there might be many reasons why individuals submit fake data, ranging from bad jokes to personal benefits by manipulating data as if quarantine restrictions were followed when they were in fact not. Also in that case, cryptographic solutions might be necessary to certify data sources while not leaking all private details. As an example, private identity servers could certify group membership in a COVID-19 risk group without leaking details about the exact medical precondition.

6 Call for Action

In order to minimize the risk of new infections and to slow down the spread of the coronavirus, a smartphone app is an effective solution, applied on an individual level. Such a COVID-19 smartphone application would enable citizens to be alerted when they are in close proximity of virus-positive tested individuals, or at high risk due to a high density of crowds in e.g. in supermarkets. It would enable high precision self-isolation advice, which is most effective to fight this pandemic.

This is a call for action to our government, to industry and to academia – for the health and safety of our population: we need to get active and develop such a COVID-19 app with technology-ensured data privacy. Additional technical support, government support and industry support is needed. We are all in this together, we should not compete but we need to collaborate!

7 apheris AI can help

Data Privacy Technology: We are a deep tech company from Berlin and are specialists in combinations of privacy preserving computations on mobile devices. We are collaborating with the world's largest open source community for privacy preserving artificial intelligence, OpenMined [21], and have the most advanced code base for on-device privacy preserving computations including access to an entire community that helps us push these technologies into production. We are dedicating all our technology and expertise to this cause.

Technical Development: We have experience working with the world's largest telecommunication and healthcare companies and can help coordinate and navigate all technical developments.

About apheris AI and the authors

We are a team of leading experts at the intersection of data privacy, cryptography and biomedical data science and work with the world's largest pharmaceutical and telecommunication companies on privacy preserving artificial intelligence.

Robin Röhm, CEO of apheris AI, studied Medicine, Philosophy and Mathematics with a focus on Cryptography

Michael Höh, PhD, CTO of apheris AI, background in physics and computer science with a focus on data science and artificial intelligence

Inken Hagestedt, PhD-researcher on privacy-preserving methods for biomedical data sharing at the Cispa Helmholtz Center for Information Security

Andrea Palladino, PhD, Physicist, working at Deutsches Elektronen-Synchrotron, COVID-19 researcher and collaborating with apheris AI

Michael Withnall, computational chemist with PhD-research on secure and private chemical data sharing working with apheris AI

Christoph Hartig, full stack engineer at apheris AI with a focus on privacy engineering

Sabrina Steinert, biomedical data scientist at apheris AI with previous experience in the pharmaceutical industry

Leonie Frauenfeld, Dr. med., medical doctor and pathologist-in-training, closely collaborating with apheris AI

ACKNOWLEDGEMENTS A.Palladino has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant No. 646623)

References

[1] Guo, Y., Cao, Q., Hong, Z. et al. The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak – an update on the status. *Military Med Res* 7, 11 (2020). https://doi.org/10.1186/s40779-020-00240-0

[2] Singhal, T. A Review of Coronavirus Disease-2019 (COVID-19). *Indian J Pediatr* 87, 281–286 (2020). https://doi.org/10.1007/s12098-020-03263-6

[3] Lai, C. C., Shih, T. P., Ko, W. C., Tang, H. J., Hsueh, P. R. (2020). Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. *Int J Antimicrob Agents*, 55(3), 105924. doi:10.1016/j.ijantimicag.2020.105924

[4] Wu, J.T., Leung, K., Bushman, M. et al. Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. *Nat Med* (2020). https://doi.org/10.1038/s41591-020-0822-7

[5] Velavan, T. P., Meyer, C. G. (2020). The COVID-19 epidemic. *Trop Med Int Health*, 25(3), 278-280. https://doi.org/10.1111/tmi.13383

[6] Li, Q., et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. New England Journal of Medicine 382, 1199-1207 (2020).

[7] Kermack, W.O., McKendrick, A.G. Contributions to the mathematical theory of epidemics I-II-II. *Bltn Mathcal Biology* 53, 33–55, 57–87, 89–118 (1991). https://doi.org/10.1007/BF02464423, https://doi.org/10.1007/BF02464424, https://doi.org/10.1007/BF02464425,

[8] https://github.com/apalladi/datascienceprojects

[9] https://www.la7.it/aggiornamenti-sul-coronavirus/video/la-virologa-ilaria-capua-23-dei-contagi-potrebbero-essere-asintomatici-25-02-2020-309481

[10] https://www.gimbe.org/pagine/341/it/comunicati-stampa

[11] https://covstat.it/

[12] https://www.uptodate.com/contents/coronavirus-disease-2019-covid-19

[13] Rhodes, A., Ferdinande, P., Flaatten, H. et al. The variability of critical care bed numbers in Europe. Intensive Care Med 38, 1647–1653 (2012). https://doi.org/10.1007/s00134-012-2627-8

[14] https://www.reuters.com/article/us-health-coronavirus/its-ok-to-feel-scared-coronavirus-testing-must-ramp-up-globally-idUSKBN2130EQ

[15] Zu, Z. Y., Jiang, M. D., Xu, P. P., Chen, W., Ni, Q. Q., Lu, G. M., Zhang, L. J. (2020). Coronavirus Disease 2019 (COVID-19): A Perspective from China. *Radiology*, 200490. https://doi.org/10.1148/radiol.2020200490

[16] Cho, H., Ippolito, D., Yu, Y. W., (2020). Contact Tracing Mobile Apps for COVID-19: Privacy Considerations and Related Trade-offs. arXiv:2003.11511 [cs.CR], https://arxiv.org/abs/2003.11511

[17] Shokri, R., Stronati, M., Song, C., Shmatikov, V. (2016). Membership Inference Attacks against Machine Learning Models. arXiv:2003.11511 [cs.CR]. https://arxiv.org/abs/1610.05820

[18] Salem, A., Zhang, Y., Humbert, M., Berrang, P., Fritz, M., Backes, M. (2018). ML-Leaks: Model and Data Independent Membership Inference Attacks and Defenses on Machine Learning Models. arXiv:1806.01246 [cs.CR]. https://arxiv.org/abs/1806.01246

[19] Dwork, C. and Roth, A. (2014), "The Algorithmic Foundations of Differential Privacy", Foundations and Trends® in Theoretical Computer Science: Vol. 9: No. 3–4, pp 211-407. http://dx.doi.org/10.1561/0400000042

[20] Pinkas B., Rosulek M., Trieu N., Yanai A. (2019) SpOT-Light: Lightweight Private Set Intersection from Sparse OT Extension. In: Boldyreva A., Micciancio D. (eds) *Advances in Cryptology – CRYPTO 2019*. CRYPTO 2019. Lecture Notes in Computer Science, vol 11694. Springer, Cham. https://doi.org/10.1007/978-3-030-26954-8_13

[21] https://blog.openmined.org/apheris-openmined-pytorch-announcement/