



The ethical use of high-performance computing and artificial intelligence: fighting COVID-19 at Barcelona Supercomputing Center

Ulises Cortés^{1,2} · Atia Cortés² · Dario Garcia-Gasulla² · Raquel Pérez-Arnal² · Sergio Álvarez-Napagao² · Enric Àlvarez¹

Received: 22 January 2021 / Accepted: 15 April 2021 / Published online: 6 May 2021
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2021

Abstract

The COVID-19 pandemic has created an extraordinary medical, economic and humanitarian emergency. Artificial intelligence, in combination with other digital technologies, is being used as a tool to support the fight against the viral pandemic that has affected the entire world since the beginning of 2020. Barcelona Supercomputing Center collaborates in the battle against the coronavirus in different areas: the application of bioinformatics for the research on the virus and its possible treatments, the use of artificial intelligence, natural language processing and big data techniques to analyse the spread and impact of the pandemic, and the use of the MareNostrum 4 supercomputer to enable massive analysis on COVID-19 data. Many of these activities have included the use of personal and sensitive data of citizens, which, even during a pandemic, should be treated and handled with care. In this work we discuss our approach based on an ethical, transparent and fair use of this information, an approach aligned with the guidelines proposed by the European Union.

Keywords COVID-19 · Artificial intelligence · Machine learning · High-performance computing · Ethics · Data protection

1 Introduction

Before the COVID-19 pandemic, the use of artificial intelligence (AI) in everyday health care delivery and administration was minimal but growing by the day. There were severe difficulties in scaling up projects and unsolved questions about the quality of health data. Also, there were questions

concerning the robustness of algorithms in the real-world applications and a policy and an essential normative void that significantly restrains institutional and human capacity to realise AI's potential and benefits.

Today AI is at the centre of increasing global competition. Due to the possible advantages that AI offers in containing the pandemic, the number of AI-based applications has increased considerably in many areas like:

1. Warnings and early alerts to citizens and decision-makers [10, 44, 64].
2. Monitoring and prediction of the pandemic [5, 46, 54, 72].
3. Data panels [43].
4. Automating aspects of diagnosis and prognosis [3, 4, 12, 45].
5. Improving vaccines, treatments and cures [9, 18, 62, 69], and
6. Support control and tracing of digital contacts [1, 56].

In reaction to the rapid spread of the COVID-19 pandemic, the Barcelona Supercomputing Center (BSC) began working on several solutions to help control and monitor the growth of the pandemic in different parts of the world, doing

✉ Ulises Cortés
ia@cs.upc.edu

Atia Cortés
atia.cortes@bsc.es

Dario Garcia-Gasulla
dario.garcia@bsc.es

Raquel Pérez-Arnal
raquel.perez@bsc.es

Sergio Álvarez-Napagao
sergio.alvarez@bsc.es

Enric Àlvarez
enric.alvarez@upc.edu

¹ Universitat Politècnica de Catalunya, Edifici Omega 205, Jordi Girona 29, 08034 Barcelona, Spain

² Barcelona Supercomputing Center, Edifici Omega 201, Jordi Girona 1 and 3, 08034 Barcelona, Spain

extensive interdisciplinary work. These activities included the use of the MareNostrum 4 supercomputer and other infrastructures at the institution for its own investigations, and open access has been given to external research teams working to fight the pandemic. BSC stores and analyses clinical data of COVID-19 patients for the conception and production of tools that support clinicians in the diagnosis and treatment of the disease.

At BSC, we conceive AI as a cross-cutting tool that is used in different ways and can play an essential role in recognising infection patterns, explaining them, and predicting them. In all these applications, work has been carried out in emergency conditions where the lack of data in the necessary quality and quantity has become evident, and those that are unreliable are available, with much noise and collected with very different methodologies. The BSC participated in nineteen projects related to COVID-19 [11].

1.1 Projects related to the analysis of the search for drugs and vaccines

The life sciences department of the BSC is working on the prediction of the structure of protein *S* and *ACE* receptors, the prediction of the structure of the virus or the discovery of drugs and vaccines that counteract it. Several projects are currently hosting research carried out at the centre on the coronavirus and its possible treatments:

1. EXSCALATE4CoV (E4C), funded by the European Commission under the H2020 program. E4C emphasises basic research and applied research to search for newer drugs [25];
2. COV20_00505 Design of antivirals for SARS, in collaboration with the Institute of Advanced Chemistry of Catalonia (IQA) and Nostrum Biodiscovery (NBD), is focused on research on antiretrovirals that can inhibit the coronavirus that causes COVID-19 and other subsequent coronaviruses. This research is funded by the COVID-19 fund of the Spanish Carlos III Health Institute [21].

1.2 Tools to analyse the spread and the medical and social impact of the pandemic

Other projects, like the global data science project (GDSP) for COVID-19 in collaboration with UNICEF and Facebook, use the available datasets with social network-based mobility data of people in Barcelona, Tokyo and New York, in a project that aims to analyse the socioeconomic impact of viruses both locally and globally, focusing on the impact of safety distance measures [68]. We also have other databases, including clinical case data, socioeconomic data, textual data, and biomedical data, that are being analysed (see Sect. 2.2).

The Hospital Clínic de Barcelona (HCB) and the BSC have worked together to create an AI-based model that helps doctors predict the evolution of patients with COVID-19 and those responsible for the centres to plan their internal organisation in the event of a new wave.

In parallel, BSC is part of the CIBERESUCICOVID consortium [66] where the objective is to analyse historical data from patients that have been treated in intensive care units (ICU) among more than thirty hospitals across Spain. AI is being used to predict the evolution of patients once they enter the ICUs, as well as to recommend treatments to medical staff.

The Computational Biology group of the Department of Life Sciences is working on the development of a publicly accessible geographic information system on the expansion of COVID-19 outbreaks (FlowMaps) [27], which integrates different data sources from public administrations, to help with the analysis of the expansion of the pandemic and decision-making related to the management of new outbreaks of COVID-19. The platform gathers data on health, mobility and geolocation from the Ministry of Health, the Ministry of Transport, Mobility and Urban Agenda, the National Institute of Statistics, the Carlos III Health Institute, the Catalan and Basque health agencies, among others. The analysis of the data is carried out with network systems that seek the relationships between them, with the aim of obtaining a better understanding of the spread of the disease. With the results, reports are made for health authorities and tools are developed to nurture epidemiological models intended to support decision-making.

PhysiBoSS-COVID [49] is an effort to integrate MaBoSS, a stochastic Boolean modelling software, into PhysiCell-COVID to allow the leverage of cell- and pathway-specific Boolean models. To obtain these COVID-19-specific models, BSC has taken advantage of CaSQ ability to convert all C19DM maps into SBML-qual files, that can subsequently be transformed to MaBoSS-format Boolean models, ready-to-use with PhysiBoSS-COVID. As a proof of concept, there has been an integration of a model of apoptosis on human epithelial host cells as a consequence of SARS-CoV-2 infection or T cell induction [48].

1.3 Plan of the work

This paper has the following structure. First, in Sect. 2, we address, from an EU point of view [34], the Ethical, Legal, Socioeconomic and Cultural issues of using AI to fight the pandemics. In Sect. 2.2 we give notice of the Global Data Science Project, and in Sect. 3 we describe the dataset and the pandemics situation in Spain. Also, we analyse the Google and Facebook datasets, for Spain, for the period from March 2020 to January 2021, and in Sect. 5 we present a general discussion and give our conclusions.

2 Ethical, legal, socioeconomic and cultural issues of using AI

Sound ethics and risk assessment processes are needed to warrant that AI is used responsibly in response to COVID-19 or any other global emergency. It is clear now that current processes for ethics and risk assessment encompassing uses of AI are still relatively immature, and the urgency of a crisis of this magnitude highlights their limitations [67].

All AI-based systems relying on the processing of personal and sensible data should be based on the principles of Convention 108+ of the Council of Europe [20], which must still be applied fully and under all circumstances regardless of whether it is the use of data about biometric data, geolocalisation, facial recognition or the use of health data. Information about socioeconomic factors like education, income, occupation, race/ethnicity have to be protected in the same way since they are considered as fundamental determinants in human life span, well-being and health [53].

Big IT companies, countries and authorities collect detailed data during outbreaks, and COVID-19 just augmented this trend. In many cases, data owners may not be able to share those datasets openly due to various ethical, legal and privacy issues, political regulations and concerns, and/or computational limitations. Besides, there are no standardised data formats that facilitate the open reporting of such information while ensuring compliance with data privacy regulations (primarily deanonymisation) [37].

As Bengio et al. [8] explain, no identifiable data should be shared with any public institution or private enterprise. Pseudonymised or aggregate data can be adequately used to develop machine learning and epidemiological models and inform public policy. Otherwise, data should be kept encrypted on users' devices and inaccessible to public authorities or private interests. A tracing application itself can propagate alerts to high-risk contacts and can recommend that users voluntarily contact health authorities where relevant, thereby assisting markedly in contact tracing while minimising the potential for state surveillance, snooping, or vigilantism. Yet, the potential gains of using personal health data to produce new knowledge cannot be minimised, for example, in the context of testing much-needed drugs and vaccines, as currently highlighted by the COVID-19 crisis [50].

We believe that robust and consensual ethical frameworks, for example [34], are needed to identify, assess and choose the correct course of action concerning the risks, opportunities and technical and moral issues posed by the use of AI. Such ethical frameworks also need to have a solid legal basis to prosecute offenders, see [22].

The design of new applications based on AI that considers in advance the ethical and legal aspects of the use of personal data is a step forward.

A data protection focus on individuals, an awareness of the social consequences of data use and the link with personality rights may expand the data controller's approach beyond data protection to fundamental rights and collective interests [42].

The use of contingency measures should be carried out in full consultation with data protection authorities and respecting the dignity and the users' private lives. How can societies navigate the grey zone of ending lockdowns while avoiding any breach of privacy and freedom? There are no easy answers to the question of the proportionality of mobility or contact-tracing technology or any of the other possible surveillance intervention. The different biases of the various types of surveillance operations should be considered, as these may cause significant discrimination [8, 23]. Our approach is to exploit data for good, avoiding any possible user identification.

2.1 Why is mobility analysis important/useful?

Since March 2020, it is clear that the COVID-19 pandemic has drastically altered mobility systems worldwide. Due to lockdowns, security distancing and stricter hygiene requirements, demand for personal mobility has gone down, while operational complexity has grown. In this time of crisis, modelling human behaviour (rational and emotional) is an essential aspect of the scenario analysis. During COVID-19, compliance with government orders like stay-at-home was one of the primary behavioural drivers to reduce contact between individuals to slow the spread of the SARS-CoV-2 and to restart economic activity [16]. For example, daily mobility data on how many kilometres were driven within each zip code in the country became a proxy for the effectiveness of the stay-at-home orders.

A relevant proportion of scenario analysis in the context of mobility is carried out with the use of AI methods, especially agent-based models and machine learning [65]. Models based in the multi-agent paradigm are used in the simulation of population mobility for the prediction of the impact of policy making [15, 17, 41]. In this kind of application, the system is usually built from the bottom up by modelling individual units (e.g., people or vehicles) under the assumption that a realistic behaviour will emerge. Another approach consists in using machine learning and data analysis algorithms, combined with high amounts of potentially heterogeneous collected data, in order to model the behaviour of the system. Such models can be applied for a myriad of application classes such as trajectory prediction [2, 71], inference of behavioural patterns [61], detection of unplanned events in urban settings

[28, 30], route planning [6], and so on. In this paper we will focus on the use of data analysis techniques for diagnosis in the context of the evolution of the pandemic.

At the beginning of the pandemic, governments and decision-makers had little precedent information for deciding the best policies to apply and to identify which measures would be the most effective for containing the virus spread while, at the same time, not restricting citizens' freedom considerably, i.e., in terms of mobility. This fact made monitoring the effect of all decisions taken a crucial point. Specifically, on the field of mobility, being able to monitor the mobility of the population, both locally and between communities, and compare it with the actual spread of the virus could help discern which mobility restrictions work better and where these restrictions are being followed adequately, even if the mobility datasets on-hand have limitations, e.g., such datasets cannot cover all population in an unbiased way [16]. In the end, this information could lead to better decisions, which could help avoid more deaths caused by COVID-19.

Therefore, there is a need for using data analysis techniques to get insights and make predictions based on scenarios of pandemic. However, in the design of any algorithm that can provide outputs consequently used for policy creation or modification, we must be aware of whatever adverse effects could they end up provoking and which possible risks are associated with them. Could mobility patterns be used for discriminating individuals or groups? Different areas might encompass different socioeconomic levels, which could be related to different mobility patterns. Is it right to use this information on law enforcement? If individuals, in a specific area, are not fulfilling the movement restrictions this should be made visible during mobility analysis. Therefore, having this analysis at hand could enable local authorities to apply local policies in a better informed way.

Before the COVID-19 pandemic, the availability of mobility data was highly restricted. Neither researchers nor public administrations could easily access them, apart from those being gathered for specific purposes like public transportation usage or survey-based origin-destination matrixes. The need for mobility analysis has pushed the big technological companies to release new mobility datasets based on, for example, Call Details Records, anonymous GPS traces and mobile application data from smartphone users. Even if sparse, the richness of these datasets and their capacity to track movement in near real time have attracted the attention of decision-makers and epidemiologists.

In Europe, on 8 April 2020, the European Commission asked European Mobile Network Operators (MNOs) to share anonymised and aggregated mobile positioning data. In compliance with the "Guidelines on the use of location data and contact tracing tools in the context of the COVID-19 outbreak" by the European Data Protection Board [24], this data does not provide information about the behaviour

of individuals; it can, however, give valuable insights into mobility patterns of population groups [59].

Therefore, data quality may vary substantially across diverse data streams. This fact was particularly evident early during the COVID-19 outbreak when data was often less structured [58]. Also, biases only become manifest after data is collected and analysed. Therefore, while researchers and decision-makers need to perform their assessments of the feasibility of specific data to support their study findings, it is compelling to have a fair data validation process in place before it is used [37].

Each company has freely decided and shared their anonymising methodologies and the data granularity they would share to avoid miss-usages of it. So, the private companies have all the decision power about the data that users generate and governments need [70]. This could raise questions about who should control mobility data.

To be ready for its ethical and transparent use, all available mobility data ought to pass through an aggregation and anonymisation process before being released to the public. Administrative regions of level 3 are used to split the data using the smaller granularity available until today. This aggregation makes it impossible to discern one specific user's mobility patterns on its own. But we must be aware that the same user could be generating data for more than one platform, which could imply several independent data handlers. There is a risk that one data handler failure could lead to this user's patterns being identified by crossing all available data. For example, people living on the same zone of a city could have clusters with specific mobility patterns if some user appears on its work website, which could help discern which of the mobility clusters this user belongs to. Additionally, low populated areas could make it easier to identify and track individuals with enough information coming from different sources.

Once we have anonymous mobility data that we can analyse, it is essential to put this data in context in order to obtain a better analysis and to be fair with all regions being analysed. Mobility patterns might be different, given each geographic area and each main economic activity. Not taking into account these two points could provide an incomplete picture, leading to sub-optimal decisions.

In the same country, there can be substantial geographic differences. This fact can have huge effects on mobility. In the late years, most rural areas have become progressively emptier as people have migrated to big cities. This implies there are some significant areas with a very sparse population. People living in this area might need to travel a considerable distance to buy something because it is not available in their village. This could mean a big data point where distance is not correlated with the number of people having contact or the infection risk. On the other hand, someone living in a city could have much more risk with half the

movement. Analysing this data without context could lead to increase the measures on rural areas guided by the considerable mobility, isolating people without a real need for it, with all the psychological effect this could have on these communities.

Mobility patterns can be strongly affected by economic activities in a specific area. In a geographical location where the main activity is related to technology, most workers can develop their work from home. This is not the case for the industry and services sectors, where workers need to be present at their workplaces. It implies that mobility restrictions will affect people's lives, in a degree that might depend on their industry. For this reason, decisions according to mobility must take into account the distribution among the economic sectors on each area and take the adequate measures to protect all the people who will not be able to work while restrictions are active.

These are two of the main concerns while analysing mobility data, but more might arise over time. People's behaviour is not one-dimensional; mobility might be related to geographic and socioeconomic factors, work sector, educational level and even psychological state of the population. Decisions on mobility restrictions might profoundly affect a community. Researchers have the responsibility of being aware that releasing an analysis about this field could have unintended effects that strongly affect our society and there must be an intent on minimising the risk of given effects being negative.

2.2 The global data science project

The main purpose of the global data science project (GDSP) is to quantitatively measure the impacts of the COVID-19 pandemic on our societies regarding people's mobility, health, and behaviour changes, and to inform public and private decision-makers to make effective and appropriate policy decisions. The GDSP team is formed by computer scientists, data analysts, designers and legal experts from institutions from several countries and is advised by members of the UNICEF Office of Innovation. Work done in GDSP is multidisciplinary, including topics such as, e.g., the impact on mental health based on collected tweets [38] or the impact on flight networks [63].

One of the main focuses of the GDSP is to analyse the seeming trade-off between economics and the prevention of infection spread. Based on the calculation of a physical distance index (mobility index), economic damages, and the number of newly infected individuals, the team analyses the optimal level where we embrace both the steady decline in the number of infections and recovery economics. While intuitively it seems clear that there is a direct relationship between physical distance and the speed of the pandemic

spread, policy makers still struggle to find the proper tools to understand the possible containment measures to apply [68].

One important subset of containment measures are those that try to affect mobility. Because the pandemic has hit worldwide, it is essential to be able to build mobility indexes that can be scaled to many countries and spanning comprehensive-time ranges. Obtaining detailed quality mobility data in a homogeneous form that tackles these issues is challenging [57]. Mobility can consist of movements varying in length and frequency: from school or work commutes to migratory movements across national borders. Typical mobility data sources such as origin-destination matrixes or GPS traces can cover some part of it (see Sect. 2.1). Still, it is hard to find data sources ready to use covering many types of mobility activities simultaneously.

For our analysis, we focus on Spain's case while aiming at being able to scale our analysis to other countries applying similar policies. In Spain we find enough elements to drive our analysis: variable pandemic spread over time and the adoption of different kinds of political measures to contain. We focus on the first wave of the pandemic, including a complete lockdown starting March 15th following mild restrictions and finishing gradually after a couple of months [32]. We studied Spain's demographics and considered the relation between restriction policies, social behaviour and pandemic evolution [29, 52]. The outcome of this study could help other countries into how to react to future similar crisis, for example in the form of adaptation of mobility policies.

3 Tracking mobility in Spain during the pandemic

Potentially, the infrastructures created in battling the pandemic may enable governments, or private companies, to exercise an immoderate amount of control over individual citizens. Therefore, any responsible implementation of mobility tracking should give particular weight to systemic risks, and ELSEC considerations [47].

Due to the aforementioned lack of data sources available in different geographical locations in a homogeneous form, we focused on analysing the feasibility of using mobility data produced by the big technological companies. Many instances of such data became available in an effort to help scientists in COVID-19 research. We studied two of these datasets, the ones published by the Facebook Data for Good program (FDG) [26] and by Google [7]. Both are based on aggregations of anonymised data generated by mobile devices.

It should be noted that by using data of this nature a certain amount of bias is expected. The penetration of mobile devices varies considerably among groups of age or different

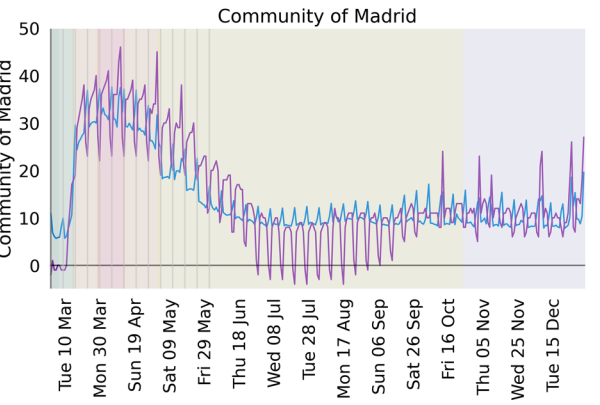
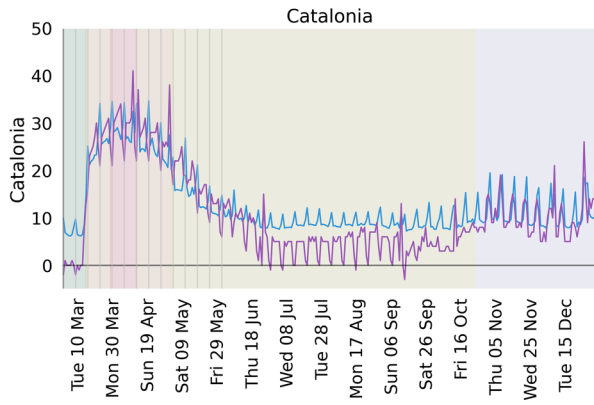
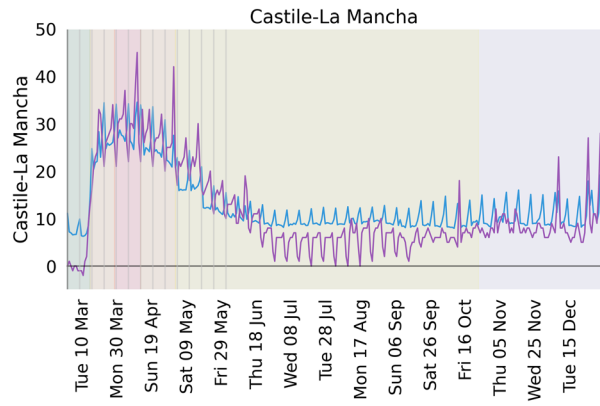
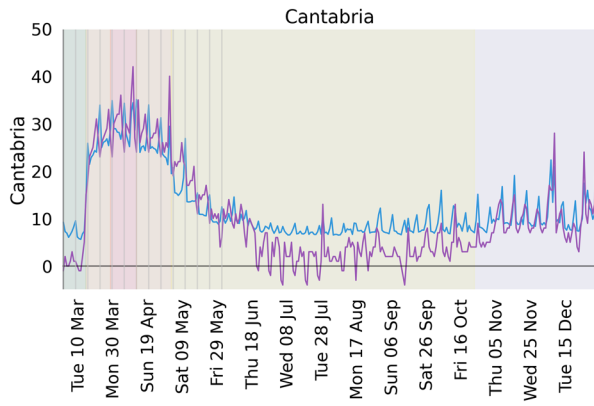
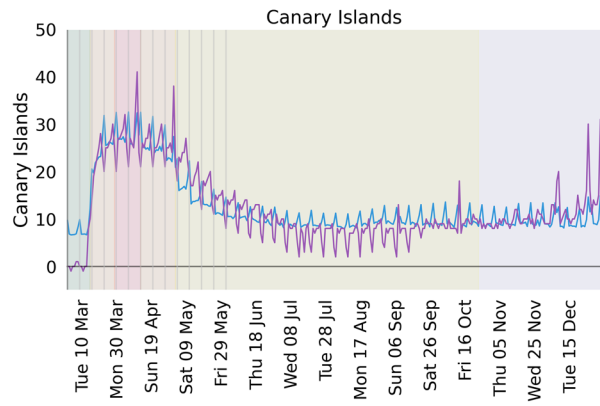
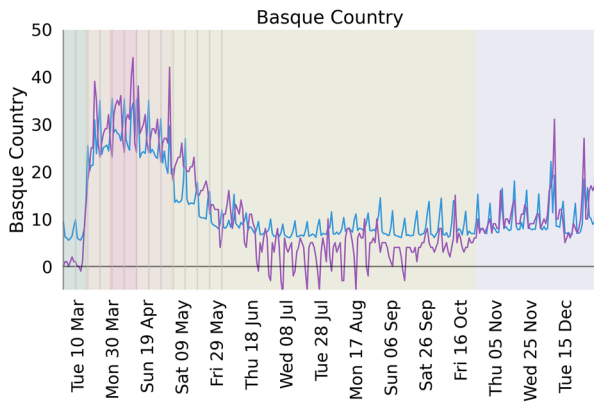
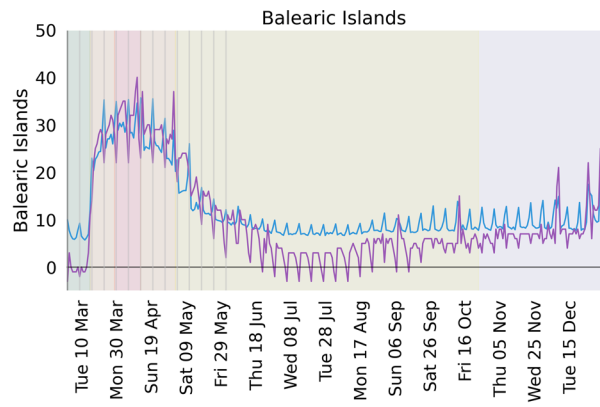
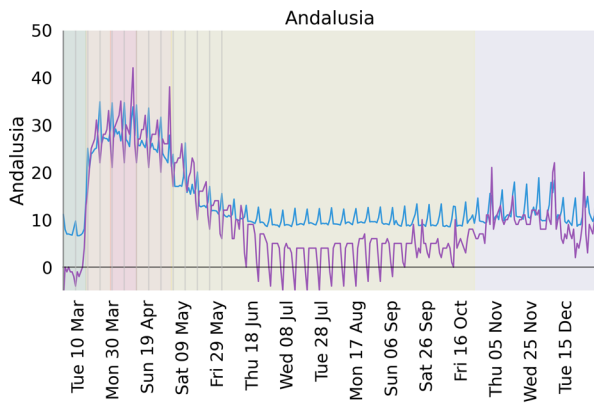


Fig. 1 Evolution of mobility between March 2020 and January 2021 in eighth autonomous communities: Facebook Remain in Tile (blue) and Google (purple), under different policy measures: no confinement (green), partial lockdown (orange), hard-lockdown (red)

socioeconomic levels. For instance, both the elders and the youngest are expected to be underrepresented.

3.1 Composition of the datasets

Usually, geographical data is aggregated based on a standard level of spatial granularity called administrative level: 1 (usually countries), 2 (regions inside countries) and 3 (administrative divisions of regions). Google data is aggregated at level 2 while FDG data is aggregated at level 3, so we aggregate the latter dataset to level 2 for our study.

Facebook Data for Good contains several metrics but two of them are of particular relevance for us: the remain-in-tile index, which is the percentage of people that remain inside the same tile of around 500×500 m during a whole day [33, 40], and the mobility index which compares the number of visits to different tiles per person any given day and compares it with the same number pre-pandemic on February 2020 providing a reduction/increase in mobility compared with that month.

Google periodically releases their COVID-19 Community Mobility Reports [7, 31]. These include several metrics corresponding to different types of activity, such as shopping or going to work. In our case we use the Residential metric, which aggregates the average number of hours spent at each user's residence.

Both metrics, FDG's remain-in-tile and Google's Residential should strongly correlate, as a considerable amount of people should score similarly in both indices at the same time individually or in aggregation. The direct comparison between both should provide context for their interpretation and insights on their applicability.

3.2 Analysing the data

Our first analysis focuses on the general mobility trend of Spanish regions during the pandemic's peak. For this, we focus on the three months around the pandemic's first wave in Spain, as shown in Fig. 1 where we show the evolution of the Google's residential index and the remain in tile index for Facebook as very good proxies of the level of confinement at home, i.e., if you remain at home you must count as a person who remained in tile and stayed a lot in a residential area.

The analysed months are March, April and May. In this period, the legal environment that affected mobility in Spain went through five clear stages. On March 1, there was a normal legislation and mobility was expected to be normal as it

can be. That would mean around zero on the Google index since this is a measure relative to normality. Later, from March 14 to March 27, there was a state of alarm, a period of mobility containment given that there was a general lockdown. Everyone had to remain at home except for shopping. No outdoor exercise or any other activity was allowed except for some strategic economic sectors. This lockdown was strengthened from March 28 to April 12 where only the most strategic sectors were allowed to function. A hard lockdown where any economic activity which required in-person work was forbidden except for essential services. This hard lockdown ends in April 13. From this point on until May 1 the more general lockdown remains in place, but reinforced policies are lifted and mobility to the workplace for certain non-essential sectors is allowed. Finally, from May 2 to June 20 (7 weeks), the legal framework became slowly heterogeneous. Different regions had a different legal framework. All of them pointed to a convergence towards new normality, as de-confinement measures were deployed. First, kids are allowed outside for a limited time, some stores are allowed to open under certain conditions, etc. This process is progressive, with new lifted restrictions every 2 weeks. It goes on until June 21, when the state of alarm ended and new normality began.

We can observe small differences in the first stages of the lockdown in different Spanish regions. However, most of them follow roughly the same path of a very sharp reduction in mobility. According to both mobility indicators, the Spanish society assumed and implemented the general lockdown in a matter of 48 h (from March 12 to March 14). Spain sustained a very high level of mobility restraint along seven weeks although it was specially so during the first four weeks, corresponding to initial lock down and then the hard lockdown (consecutive orange and redbands in Fig. 1). It kept this reduction rather high for another two-three weeks and, later on, mobility started to come back to normal. There are some differences in the level of confinement and speed of recovery, which can be explained by differences in the speed of de-confinement applied by each region.

3.3 Expected insights of mobility patterns and COVID-19 in Spain

Technological companies developed the mobility indexes with the aim in mind to help governments and citizens to understand the interplay between legislation, mobility and epidemics. Personal data were not publicly available and, indeed, it can be used to understand better the epidemics in Spain.

To observe the effect and timing of the policies implemented, we can compare it with the pandemic status. For that purpose, we use the number of new hospitalisations

plotting it against the mobility curves. To approximate the contagion date from the report's date, we shift this data two weeks early.

It is important to understand why one should use new hospitalisations to compare mobility data and the epidemics. During the first wave of the epidemics last March, most cases in Spain were detected at the hospital. Tests were not deployed in Primary care and there was no contact tracing. Therefore, the National Health System only detected a small fraction of the cases, estimated at around 10% [13]. On the other hand, cases at that time were not properly assigned the day of symptom and time series are highly unreliable.

So, new hospitalisation was the most reliable data, together with reported deaths in hospitals, not because all deaths from COVID-19 were properly reported but because they were a consistent metric of the epidemics evolution. However, we must notice that there is a time lapse between becoming infected, developing symptoms and, in case the prognosis of the disease worsens, hospitalisation.

This shift is motivated by current estimates where 4–7 days lapse between infection and another week or more until prognosis worsens. A 10–14 days shifts are reasonable to estimate how mobility can affect the course of hospitalisation thanks to the reduction in infections. Figure 2 shows this comparison, for the case of Madrid [14], the region with the most cases and the strongest lockdown adherence. We

observe the beginning of the lockdown overlapping with the initial containment of the pandemic. Seven weeks of general lockdown coincide with the 7 weeks of most substantial pandemic rate reduction. That is when it took the pandemic to reach a basal situation in the region, with a very low level of hospitalisations.

We must stress a rather important point regarding the lifting of measures. During the crisis, policy makers could only use hospitalisation data as a reliable source of information. Actually, daily cases presented a very short delay compared with hospitals, indicating precisely that COVID-19 was detected precisely around hospitals. Both data series presented, basically, the same information. According to this estimate, a proper control of the epidemics was actually achieved earlier than the hospital data suggest. It is perfectly possible that the Spanish lockdown duration (7 weeks) was a very good fit for the pandemic's evolution.

Despite the possibility that the epidemics was under control earlier than hospital data suggested, a shorter lockdown may have induced a significantly higher risk of relapse. Keeping mobility down for more weeks helped to push the real number of infection down. And still, these numbers were not low enough to guarantee strong suppression of the virus. Actually, as shown in Fig. 1, cases started raising again in Catalonia during the summer, only a couple of months after the end of the lockdown. If sustained for

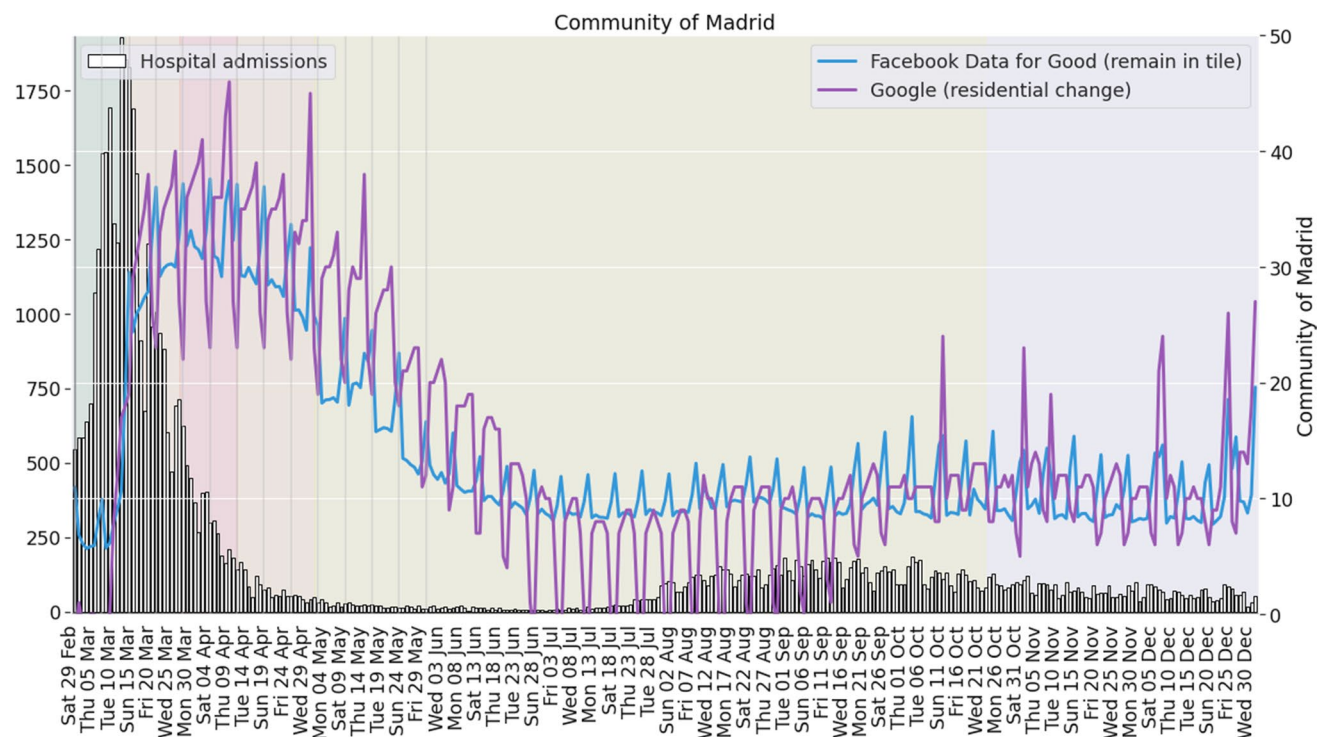


Fig. 2 According to Facebook's remain-in-tile and Google's residential index, mobility containment in Madrid, together with the number of reported hospital admissions daily in Madrid [14] shifted 10 days early to approximate contagion date

longer, it would certainly have driven the number of cases to very low levels. Therefore, the de-confinement process was a bolder (and riskier) initiative than Fig. 2 illustrates.

The role of the hard-lockdown versus regular lockdown is also a source of knowledge for future epidemics control. The regular lockdown includes five weeks of lockdown data (the last two weeks of March and the first three weeks of April) during which travelling to industry and construction workplaces was allowed. This sort of lockdown is assumed to have a less damaging effect on the economy, but it enables co-workers' infection. To compare mobility between both periods, we measure the corresponding area under the curve, which indicates an increase between 10 and 15% in the people who stayed put. The relevance of that number for the containment of the pandemic is unknown to us. i.e., we do not know which would have been the pandemic's evolution if hard-lockdown had not been implemented. One may argue that many more people would have been infected since regular lockdown enables transmissions on the workplace.

However, the number of detected cases did not show any change after hard-lockdown ended; it continued to decrease at a similar rate. One may also argue that the hard-lockdown had a psychological effect on society, boosting resiliency to confinement. Looking at how mobility recovers right after hard-lockdown is lifted, this seems to be a valid hypothesis. If that were the case, a state of the alarm without hard-lockdown might have lost adherence faster.

4 Ethical aspects of non-expected mobility patterns insights

We address now a key question that arises when deploying the information of mobility data. Namely, the fact that when different sources are available, cross analysis can unveil patterns and analysis that were not initially in the mind of those developing the mobility data.

While personal data is not available and it is impossible to correlate any particular movement to any particular person, general trends and population behaviour can become uncovered. We present here two examples of this indirect knowledge. Let us start with the one that does not present any particular ethical problem since it is basically an economic-related insight that can already be obtained through other indirect measures.

Given that within the seven lockdown weeks, two were under hard-lockdown and police strongly enforced restrictions, by analysing the mobility in these two weeks we can obtain a reasonable estimate of the maximum mobility restriction that can be held in Spain while keeping essential services running. Actually, by knowing the number of non-essential services that were allowed during the

soft-confinement one can estimate the impact on mobility of each one of these sectors.

The fact that different mobility sources are available can provide useful econometric data to understand how economic policies can affect economic structure which, at the same time, can affect the type of mobility network (public transports, roads, trains..) needed to sustain it. In this case, having different sources only provides more accurate information about issues that are already being investigated by countries all around the world. In other words, it represents a spill-over of the information with no particular ethical concerns.

Let us proceed now to address a more complex example of unexpected insight from cross-analysis of data: average personal behaviour of the citizenship.

4.1 Unveiling average personal behaviour

The analysis of mobility relations between different sources of mobility data can provide insights on a very different matter. It can help to unveil average behavioural patterns of the population. While companies right now can obtain average behaviour data on different features of on-line interactions such as browse history or type of interactions, the key goal of these companies is to gather different average information from different sources so that they can obtain information about average patterns.

One must not mix this ethical question with the importance of permission structure regarding privacy. Indeed, people can opt in to let companies use personal history to direct advertisement. They can even leave details about their personal mobility. So, in a certain way, average behaviour can be obtained regarding interaction patterns if enough people give permissions to track their mobility.

Here, the question arises where information from different sources regarding mobility can be inspected, which changes in population patterns can be detected, and what are its ethical ramifications. A given company can detect from those users allowing the use personal tracking that a particular geographical spot (a commercial shopping area, or a restaurant) became more trendy. Similarly, it can detect if the number of people doing sports in gyms or any other specific activity increased. However, any of these indicators can also be obtained indirectly via browsing history. Still, there is a difference between inferring and knowing these changes.

In any case, a particular company cannot be sure that this changes in profile are unique to their customer base or are more general to the whole population. The publication of multiple average mobility data can unveil precisely these general patterns. We proceed to show that with the combination of Facebook data and its Mobility index and

Google indicators of mobility, one can observe shift in average behaviour and, more specifically, in changes in daily or weekend behaviour.

4.2 Daily Trend

Weekdays have an essential role in the characterisation of mobility. Let us now study the same data, but this time with a daily perspective. The plots in Fig. 3 show the reduction in mobility using Facebook data and Google mobility index. For each month, we provide two visualisations of the evolution of these indexes during the pandemic in Spain. The cloud of points shows how the relation between both changes strongly depending on the month. Each month had a different epidemiological situation and, as expected, different average general behaviour of the population.

On the top row, the colour gradient shows the change through time, week by week. On the bottom row, weekdays are colour coded to illustrate the differences between days. The reduction in mobility detected by Facebook is higher when the dot is more to the right. The vertical axis shows the relative change in the number of people that stay in residential areas. When this value is high, a high percentage of people stay at home with respect to regular instances of that day of the week.

In Fig. 3, the first noticeable thing is the correlation between both values during lockdown, as all data is mostly gathered around the diagonal. The top row shows the evolution of mobility, starting from the axis origin (bottom left) and suddenly jumping to the plot's top-right quarter as lockdown is implemented. The last Friday and Saturday before the lockdown (second week of March) are the only days in the middle of that jump. During the lockdown, in April, data is relatively stable in that area, until the de-confinement measures, in May, bring it down and left again, but this time in a slow manner.

On the other hand, as the number of cases decreases, mobility restriction are lifted, the cloud of points shifts completely from the diagonal. During summer there was a broad range of changes in the Google's residential index while a shorter reduction in mobility is detected by Facebook. Interestingly enough, these high correlations come back in November when the second wave of the epidemics reaches Spain, and new and softer restrictions measures were applied.

4.3 The new normality

In the previous section, we have seen the relationship between the two indexes. Here we can see how the population's average behaviour shifted as the government lifted confinement measures in the second half of May. We must

recall here again that the process was asymmetrical, with regions with better pandemic indicators (i.e., number of daily cases, number of available hospital beds, etc.) de-confining faster than others

Governmental sources offer detailed maps of the differential treatment of regions can in [55]. This process ended on June 21st, when the state of alarm and all mobility restrictions ceased. On that date, the whole country officially entered the *new* normality.

Figure 4 includes the period from the last 4 weeks of the state of alarm, without a generalised lockdown, until the end of August, covering several weeks of new normality. To facilitate visualisation, we change zoom in and the axis scale with respect to Fig. 3.

The progression of mobility towards the axes origin is visible as weeks go by (in colour gradient), for both working days (Monday to Friday) and weekend days (Saturday and Sunday). To compare the *new* normality with the old one, we must focus on the Google axis, since this is relative to a baseline (January 3 to February 6). On the weekends, mobility is already at Google baseline levels, with all values between -5 and 5 on the last week, the new normal one. In contrast, working days show a higher difference concerning the baseline, with several values in the last week between 5 and 15 in the Google axis.

This behaviour indicates that the Spanish society's change, during the new normality, is focused on working days, while weekends are back to how they were. We can analyse these indicators of change in behavioural pattern for different regions focusing on these days, where we can see that the recovery of old normality is not homogeneous. Catalonia and Madrid, the regions with the largest metropolitan areas converge into higher reductions of mobility in weekdays, compared to the regions of Asturias, Navarre, La Rioja, Region of Murcia, Extremadura and Galicia where the restriction measures had a lower impact on mobility.

This is likely related to the fact that Catalonia and Madrid were slightly behind in removing restriction measures during that period. Alternatively, as indicated in [29], this may also be related to the role of large metropolitan areas, where it is harder to keep a safe distance. Both regions reported the highest absolute volume of infections during the pandemic. Both of these factors are strong psychological enablers of self-responsibility, which may affect adherence to mobility reduction during the new normality. Another possibility is that these effects can be more noticeable in areas where the potential for social activity and interactions is higher.

4.4 An ethical dilemma

From the analysis made in the previous subsections, we observe that it is possible to detect the changes in the population's average behaviour with the cross-validation of

Fig. 3 The figure shows the evolution of reduction in mobility from Facebook (horizontal axis) and Google (vertical axis) data sources. Each dot represents a single day in a single region. The first column holds data for March/June/September, the second for April/July/October and the third for May/August/November. The odd rows show the data with a colour gradient indicating the week of the month. The even rows show the same data coloured based on the day of the week

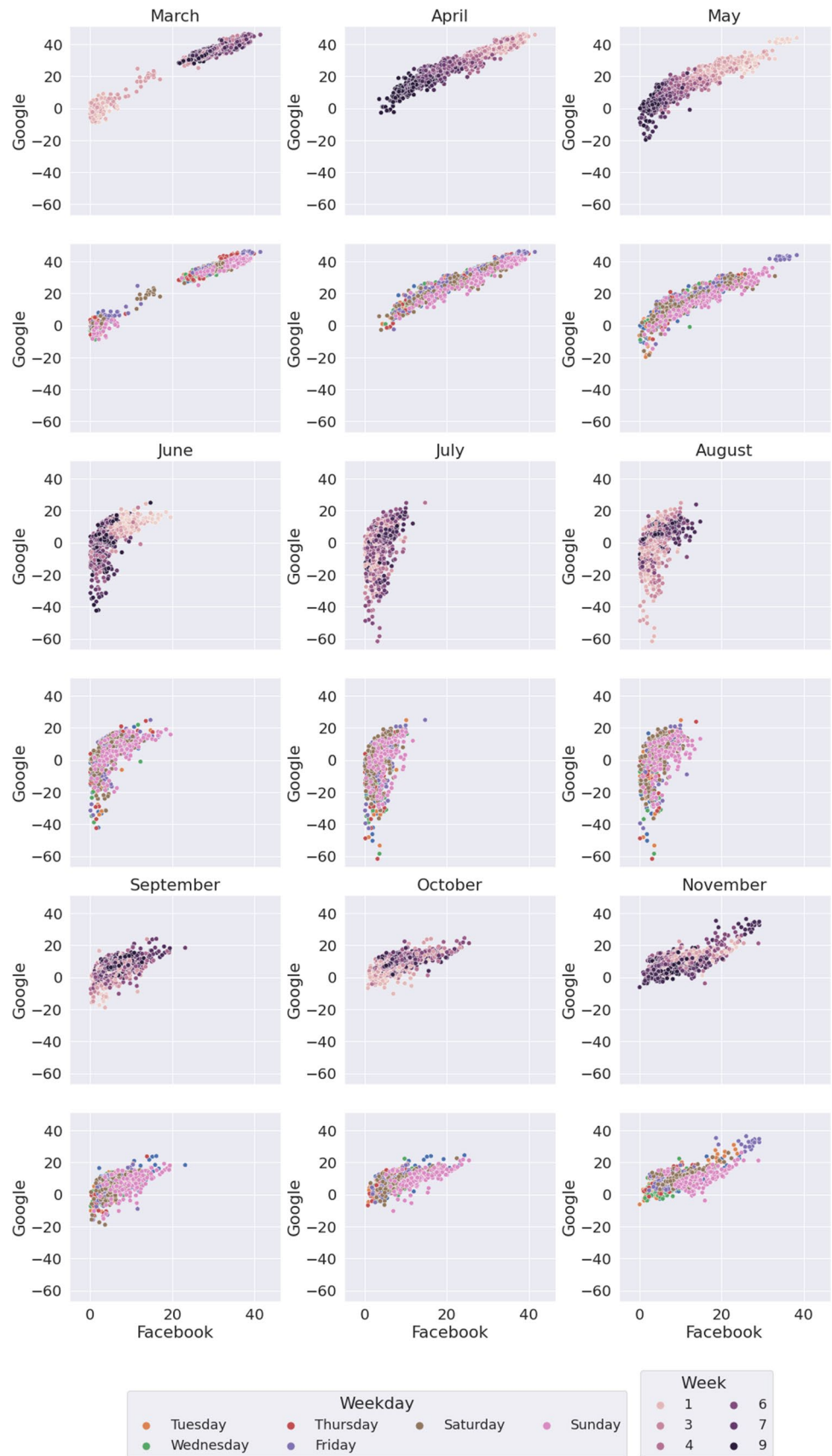
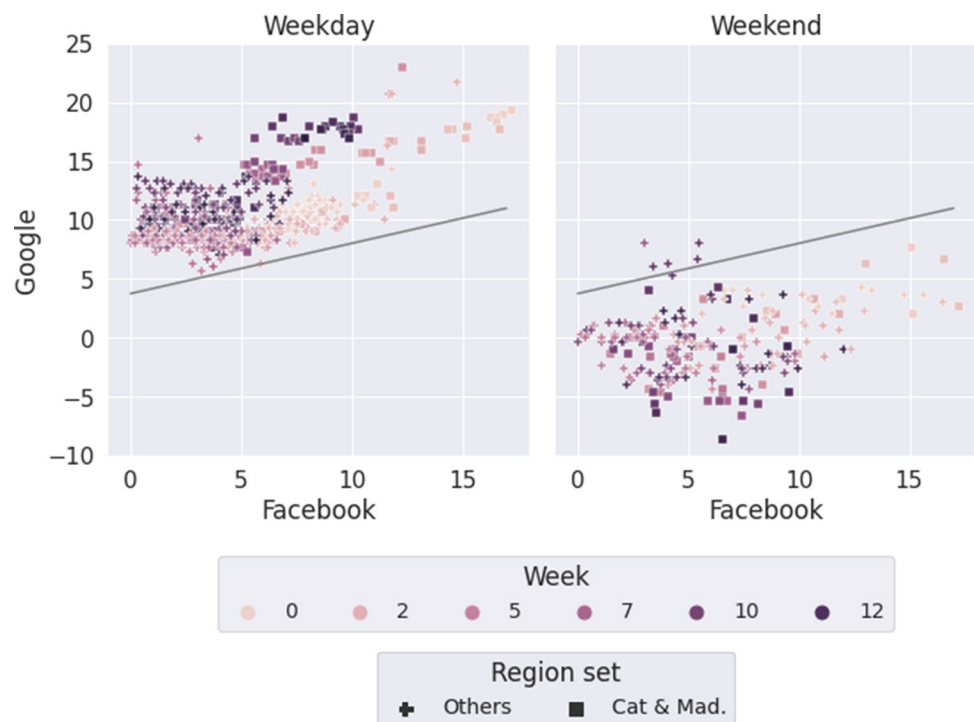


Fig. 4 This figure shows mobility values from Monday to Friday (left plot) and from Saturday to Sunday (right plot). Colour gradient indicates the week number (from May 25th to August 31st). The marker's shape indicates the region set (Crosses: Asturias, Navarre, La Rioja, Region of Murcia, Extremadura and Galicia. Squares: Catalonia and Madrid). Grey diagonal line is the same on both plots, used for visual reference



different data sources. Our research can unveil knowledge regarding general patterns. The ability of governments and/or society to be self-aware of these changes presents essential ethical questions. Is it any different than knowing average behavioural for a given particular activity that becomes more trending? What are the possible consequences regarding human behaviour knowledge? One can envision governments or companies feeding different strategies for providing information and detecting what works best. This is the type of experiment that political campaigns use to check what kind of e-mail or personal interaction may help more on fundraising in the US [60].

In any case, the possibility of obtaining information on changing patterns for the epidemics is just a small example of the famous problem of getting unrelated information from original sources, not originally designed to provide them. Still, data process is often taken-for-granted and the low quality in data may lead to poor results [58].

Another ethical aspect to take into account is the responsibility of adequately storing and protecting mobility data. Although the study presented here is purely based on anonymised, aggregated data, it is essential to have complete awareness about the nature of the data at all times. For example, privacy risks have been detected in datasets such as those coming from contact tracing apps [36] due to excessive personal information disclosure. Data such as origins, destinations, means of transport, social network data, among others, can be considered personal information under certain circumstances [19]. In general, there is a

strong public concern regarding the use of data originated in mobile devices due to the risk of damaging privacy and civil liberties [51].

We followed a preemptive measure to double-check with the providers that they understand these risks and do any pre-processing necessary to mitigate these risks. For example, we verified that Facebook—the data of which had a lower granularity—was aggregating at a scale enough to make sure that individuals could not be traced, especially regarding areas with lower population density.

5 Discussion and Conclusions

After almost one year of pandemic, digital technologies, including AI and information technologies, are demonstrating to be crucial tools to help build a globally coordinated response. Times of crisis show that societies can necessitate coordinated scientific and technological teams to the rapid deployment of new technologies to save lives and restore normality. The multiple uses of AI-based tools also put forward the limits of what can currently be achieved by these technologies, see Sects. 1.1, 1.2, which we cannot expect to compensate for structural challenges such as those experienced by many health care institutions around the world.

Simultaneously, there is growing concern that the temporary restrictions that digital surveillance entails could lead to a long-lasting suspension of civil rights and liberties and could have some unintended outcomes. In the case of AI,

COVID-19 has acted as a booster and amplified potential research opportunities and concerns.

In particular, this situation consolidated the adoption and use of AI-based techniques in scientific and medical research. This is especially evident in the field of pharmaceuticals. Never before has AI, thanks to the emergence of specific hardware, new algorithms, and the availability of large data sets, have shown the ability of machines to solve intricate biological puzzles faster than human experts. This crisis accelerated data sharing practices between the industrial sector and governments. Academia was also very active in this interchange.

Other fields where researchers and governments have shown interest are citizen's mobility and, in particular, digital contact-tracing apps [39].

5.1 GDSP takeaways

The dynamics of infectious diseases like COVID-19 are affected by human mobility more strongly than previously thought, and therefore reliable mobility traceability data is essential. To study the change in mobility, evaluate mobility restriction policies' efficiency, and facilitate better response to possible future crises, we need to understand all mobility data sources at our disposal correctly.

One of the main issues when working with private data sources is imperfect knowledge regarding its nature. For privacy-preserving reasons, raw data is never provided. Instead, data goes through heavy pre-processing and anonymization procedures [33].

In the GDSP pilot that we explained in Sect. 3, we are sure that the granular non-identifying information used to train machine-learning models do not contain sufficient detail to re-identify individuals when correlated with other sources of data [23]. Our results suggest that the data collected during the GSPD pilot could provide unmatched individualised human movement information and address significant gaps in currently-available data for mobility during and after lockdown.

In Sect. 3.3 we show that there exist significant differences among days. During the studied period, weekends exhibit the highest volume of mobility reduction in absolute terms. Even during the hard-lockdown, when travelling to work was forbidden for all except essential services. At the same time, weekends have the smallest mobility reduction in relative terms, indicating that society's effort to make in this regard with respect to its previous patterns was smaller. Fridays and Sundays are particularly relevant days. The former is because it represents the most significant deviation from normal behaviour, while the latter represents the most significant absolute decrease in mobility. Decision-makers can exploit these particularities for the general good. As Hildebrandt states, technologies can be designed in such a

way that values, ethical norms, and legal principles are built in, design for values can act as powerful limitations for possible excesses, for example, in times of crisis [35].

In the case of mobility and contact-tracing apps, citizens deserve clarity and simplicity about all aspects of their implementation. Among others: the purpose of data collection, the types of data collected, the parties who have access to them, the extent, modalities, and the timeline for data deletion, the algorithms and data training sets that will automate processes and influence their daily lives. Also, it is essential to know who is the authority in charge. Therefore it is clear that transparency is crucial for the adoption of these technologies and to maintain legitimacy and trust.

5.2 Final Remarks

In health crises like this, decision-makers must protect public health by facilitating mobility that permits citizens to meet their basic needs safely. Decision-makers, for example at City level, using the information and methods like the one described in §3, can lead the development and evaluation of strategies through social, tactical and technological policies to perform realistic interventions, some of those may remain in place after the crisis.

There is a general agreement that digital technologies, and AI-based technologies, may prove useful tools when exiting national lockdowns, as shown in Sect. 4.3, in the current COVID-19 crisis.

In light of the results presented here, it is clear that governments and scientists should collect and aggregate many types of data in addition to the studies that are currently ongoing addressing the epidemiology and biology of COVID-19. Given their potential to threaten privacy and individual liberty and foster inequalities, robust oversight of the deployment of these AI-based surveillance technologies, which involves users and civil society groups, is urgently needed [23, 39]. As Kraemer et al. indicate, trust is one of the critical components enabling rapid and efficient data sharing. The misuse of data has been detrimental to data sharing and disincentivised open collaborations [37].

The risks of unintended and negative consequences associated with AI-based technologies are commensurately high, particularly at this pandemic scale. Questions remain regarding how digital AI-based surveillance tools, as mobility tracking apps, designed for pandemic containment, will work in practice.

Acknowledgements We want to thank Facebook and Google for releasing the data that made this work possible. We also appreciate the insights and support of Ama Herdågdelen, Alex Pompe and Alex Dow on the interpretation of peaks. Part of this work was done under the Global Data Science Project for COVID-19. We would also like to thank Daniel López-Codina, Sergio Alonso, and Clara Prats for fruitful discussions.

Declarations

Conflict of interest XX belongs to a research group with a partnership agreement with Facebook Data for Good for purely research purposes. The partnership does not involve any financial or monetary relationship of any kind. The rest of the authors certify that they have NO affiliations with or involvement in any organisation or entity with any financial interest, or non-financial interest in the subject matter or materials discussed in this manuscript.

References

1. Agbehadj, I.E., Awuzie, B.O., Ngowi, A.B., Millham, R.C.: Review of big data analytics, artificial intelligence and nature-inspired computing models towards accurate detection of covid-19 pandemic cases and contact tracing. *Int. J. Environ. Res. Public Health* **17**(15), 5330 (2020)
2. Alahi, A., Goel, K., Ramanathan, V., Robicquet, A., Fei-Fei, L., Savarese, S.: Social lstm: Human trajectory prediction in crowded spaces. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 961–971 (2016)
3. Alimadadi, A., Aryal, S., Manandhar, I., Munroe, P.B., Joe, B., Cheng, X.: Artificial intelligence and machine learning to fight Covid-19 (2020)
4. Ardakani, A.A., Kanafi, A.R., Acharya, U.R., Khadem, N., Mohammadi, A.: Application of deep learning technique to manage covid-19 in routine clinical practice using CT images: results of 10 convolutional neural networks. *Comput. Biol. Med.* **121**, (2020)
5. Arenas, A., Cota, W., Gómez-Gardenes, J., Gómez, S., Granell, C., Matamalas, J.T., Soriano-Panos, D., Steinegger, B.: A mathematical model for the spatiotemporal epidemic spreading of covid19. *MedRxiv* (2020)
6. Bast, H., Delling, D., Goldberg, A., Müller-Hannemann, M., Pajor, T., Sanders, P., Wagner, D., Werneck, R.F.: *Route planning in transportation networks*. In: *Algorithm Engineering*, pp. 19–80. Springer, New York (2016)
7. Bavadekar, S., Cossoul, G., Davis, J., Desfontaines, D., Fabrikant, A., Gabrilovich, E., Gadepalli, K., Gipson, B., Guevara, M., Kamath, C., Kansal, M., Lange, A., Mandayam, C., Oplinger, A., Pluntke, C., Roessler, T., Schlosberg, A., Shekel, T., Vispute, S., Vu, M., Wellenius, G., Williams, B., Wilson, R.J.: Google COVID-19 community mobility reports: Anonymization process description (version 1.0) (2020). [arXiv:2004.04145](https://arxiv.org/abs/2004.04145)
8. Bengio, Y., Janda, R., Yu, Y.W., Ippolito, D., Jarvie, M., Pilat, D., Struck, B., Krastev, S., Sharmac, A.: The need for privacy with public digital contact tracing during the COVID-19 pandemic. *Lancet* **2**(7), e342–e344 (2020). [https://doi.org/10.1016/S2589-7500\(20\)30133-3](https://doi.org/10.1016/S2589-7500(20)30133-3)
9. Bobdey, S., Ray, S., et al.: Going viral-Covid-19 impact assessment: a perspective beyond clinical practice. *J. Mar. Med. Soc.* **22**(1), 9 (2020)
10. Bragazzi, N.L., Dai, H., Damiani, G., Behzadifar, M., Martini, M., Wu, J.: How big data and artificial intelligence can help better manage the Covid-19 pandemic. *Int. J. Environ. Res. Public Health* **17**(9), 3176 (2020)
11. BSC: BSC uses bioinformatics, artificial intelligence and the computing power of the MareNostrum supercomputer in the fight against the coronavirus (2019). <https://www.bsc.es/news/bsc-news/bsc-uses-bioinformatics-artificial-intelligence-and-the-computing-power-the-marenostrum>
12. Bullock, J., Luccioni, A., Pham, K.H., Lam, C.S.N., Luengo-Oroz, M.: Mapping the landscape of artificial intelligence applications against covid-19. *J. Artif. Intell. Res.* **69**, 807–845 (2020)
13. Català, M., Pino, D., Marchena, M., Palacios, P., Urdiales, T., Cardona, P.J., Alonso, S., López-Codina, D., Prats, C., Alvarez-Lacalle, E.: Robust estimation of diagnostic rate and real incidence of covid-19 for european policymakers. *PLOS One* **16**(1), 1–26 (2021). <https://doi.org/10.1371/journal.pone.0243701>
14. Centro Nacional de Epidemiología, Instituto de Salud Carlos III: Situación y evolución de la pandemia de COVID-19 en españa (2020). <https://cnecovid.isciii.es/covid19/#documentaci%C3%B3n-y-datos>
15. Čertický, M., Drchal, J., Cuchý, M., Jakob, M.: Fully agent-based simulation model of multimodal mobility in european cities. In: *2015 International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, pp. 229–236. IEEE (2015)
16. Chang, S., Pierson, E., Koh, P., Gerardin, J., Redbird, B., Grusky, D., Leskovec, J.: Mobility network models of COVID-19 explain inequities and inform reopening. *Nature* **589**, 82–87 (2021). <https://doi.org/10.1038/s41586-020-2923-3>
17. Chen, B., Cheng, H.H.: A review of the applications of agent technology in traffic and transportation systems. *IEEE Trans. Intell. Trans. Syst.* **11**(2), 485–497 (2010)
18. Chen, S., Yang, J., Yang, W., Wang, C., Bärnighausen, T.: Covid-19 control in china during mass population movements at new year. *Lancet* **395**(10226), 764–766 (2020)
19. Cottrill, C.D.: Maas surveillance: privacy considerations in mobility as a service. *Transp. Res. Part A Policy Pract.* **131**, 50–57 (2020)
20. Council of Europe: The Convention for the protection of Individuals with regard to Automatic Processing of Personal Data (2018). <https://www.coe.int/en/web/data-protection/convention108-and-protocol>
21. COV2/00050: Diseño de antivirales para SARA basados en polifarmacología (2020). <https://www.eu-isciii.es/wp-content/uploads/2020/05/0505.pdf>
22. European Commission: Regulation (EU) 2016/679: General Data Protection Regulation (GDPR). EC (2016)
23. European Commission: Commission Recommendation (EU) 2020/518 of 8 April 2020 on a common union toolbox for the use of technology and data to combat and exit from the COVID19 crisis, in particular concerning mobile applications and the use of anonymised mobility data. *Off. J. Eur. Union* (L 114/7) (2020). URL <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32020H0518&from=EN>
24. European Data Protection Board: Guidelines 04/2020 on the use of location data and contact tracing tools in the context of the (covid-19) outbreak (2020). https://edpb.europa.eu/our-work-tools/our-documents/linee-guida/guidelines-042020-use-location-data-and-contact-tracing_en
25. EXSCALATE4COV: EXaScale smArt pLatform Against paThogEns for Corona Virus (2020). <https://www.exscalate4cov.eu>
26. Facebook: Facebook data for good public datasets (2020). <https://dataforgood.fb.com/>
27. Ferrari, A., Santus, E., Cirillo, D., Ponce-de Leon, M., Marino, N., Ferretti, M.T., Chadha, A.S., Mavridis, N., Valencia, A.: Reproducing sars-cov-2 epidemics by region-specific variables and modeling contact tracing app containment. *medRxiv* (2020)
28. Gabrielli, L., Rinzivillo, S., Ronzano, F., Villatoro, D.: From tweets to semantic trajectories: mining anomalous urban mobility patterns. In: *International Workshop on Citizen in Sensor Networks*, pp. 26–35. Springer, New York (2013)
29. Garcia-Gasulla, D., Álvarez-Napagao, S., Li, I., Maruyama, H., Kanezashi, H., Pérez-Arnal, R., Miyoshi, K., Ishii, E., Suzuki, K., Shiba, S., Kurokawa, M., Kanzawa, Y., Nakagawa, N., Hanai, M.,

- Li, Y., Li, T.: Global data science project for Covid-19 summary report. CoRR (2020). [arXiv:2006.05573](https://arxiv.org/abs/2006.05573)
30. Garcia Gasulla, D., Álvarez Napagao, S., Tejada Gómez, J.A., Oliva Felipe, L.J., Gómez Sebastià, I., Béjar Alonso, J., Vázquez Salceda, J.: Social network data analysis for event detection. In: ECAI 2014: 21st European Conference on Artificial Intelligence: 18–22 august 2014, Prague, Czech Republic: proceedings, pp. 1009–1010. IOS Press (2014)
 31. Google: COVID-19 Community Mobility Reports (2019). <http://google.com/covid19/mobility>
 32. Hegedus, A., Annunziato, A., Gerhardinger, A., Wania, A., Delipetrev, B., Gasparro, C., Fonio, C., Proietti, C., Turk, D., Sabo, F., Rios, F., Eklund, G., Joubert-Boitat, I., Monster, J., Swen, J., Kamberaj, J., Poljansek, K., Brzostowska, K., Mastro-nunzio, M., Santini, M., Halkia, M., Kalas, M., Ferrer, M.M., McCormick, N., Probst, P., Rufolo, P., Barbosa, P., Vojnović, P., Spruyt, P., Boskovic, S.G., Kemper, T., Antofie, T., Harmatha, T., Durrant, T., Salvitti, V.: ECML covid dashboard. European Commission Joint Research Centre - ISPR - Space, Security and Migration Directorate (JRC) (2020). <https://covid-statistics.jrc.ec.europa.eu/>
 33. Herdağdelen, A., Dow, A., State, B., Mohassel, P., Pompe, A.: Protecting privacy in Facebook mobility data during the COVID-19 response (2020). <https://bit.ly/2W0ivDw>. Facebook Research
 34. High-Level Expert Group: Ethic Guidelines for Trustworthy AI. European Union, Brussels (2019). https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419
 35. Hildebrandt, M., Tielemans, L.: Data protection by design and technology neutral law. *Comput. Law Secur. Rev.* **29**(5), 509–521 (2013). <https://doi.org/10.1016/j.clsr.2013.07.004>
 36. Jung, G., Lee, H., Kim, A., Lee, U.: Too much information: assessing privacy risks of contact trace data disclosure on people with covid-19 in South Korea. *Front. Public Health* **8** (2020)
 37. Kraemer, M.U.G., Scarpino, S.V., Marivate, V., Gutierrez, B., Xu, B., Lee, G., Hawkins, J.B., Rivers, C., Pigott, D.M., Katz, R., Brownstein, J.S.: Data curation during a pandemic and lessons learned from (covid-19). *Nat. Comput. Sci.* **1**, 9–10 (2021). <https://doi.org/10.1038/s43588-020-00015-6>
 38. Li, I., Li, Y., Li, T., Alvarez-Napagao, S., Garcia, D.: What are we depressed about when we talk about Covid19: Mental health analysis on tweets using natural language processing (2020). [arXiv:2004.10899](https://arxiv.org/abs/2004.10899) (2020)
 39. Lucivero, F., Hallowell, N., Johnson, S., Prainsack, B., Samuel, G., Sharon, T.: COVID-19 and contact tracing apps: Ethical challenges for a social experiment on a global scale. *Bioeth. Inquiry* **17**, 835–839 (2020). <https://doi.org/10.1007/s11673-020-10016-9>
 40. Maas, P., Iyer, S., Gros, A., Park, W., McGorman, L., Nayak, C., Dow, P.A.: Facebook disaster maps: Aggregate insights for crisis response & recovery. In: KDD '19: Proceedings of the 25th ACM SIGKDD Int. Conf. on Knowledge Discovery & Data Mining. ACM (2019). <https://doi.org/10.1145/3292500.3340412>
 41. Maggi, E., Vallino, E.: Understanding urban mobility and the impact of public policies: the role of the agent-based models. *Res. Transp. Econ.* **55**, 50–59 (2016)
 42. Mantelero, A.: Artificial Intelligence and data protection: Challenges and possible remedies. Tech. rep., Council of Europe (2019). <https://rm.coe.int/2018-lignes-directrices-sur-l-intelligence-artificielle-et-la-protecti/168098e1b7>
 43. Marvel, S.W., House, J.S., Wheeler, M., Song, K., Zhou, Y.H., Wright, F.A., Chiu, W.A., Rusyn, I., Motsinger-Reif, A., Reif, D.M.: The covid-19 pandemic vulnerability index (pvi) dashboard: Monitoring county-level vulnerability using visualization, statistical modeling, and machine learning. *Environ. Health Perspect.* **129**(1), (2021)
 44. Mehraeen, M., Dadkhah, M., Mehraeen, A.: Investigating the capabilities of information technologies to support policymaking in Covid-19 crisis management; a systematic review and expert opinions. *Eur. J. Clin. Investig.* **50**(11), (2020)
 45. Mei, X., Lee, H.C., Diao, K.Y., Huang, M., Lin, B., Liu, C., Xie, Z., Ma, Y., Robson, P.M., Chung, M., et al.: Artificial intelligence-enabled rapid diagnosis of patients with covid-19. *Nat. Med.* **26**(8), 1224–1228 (2020)
 46. Mohamadou, Y., Halidou, A., Kapen, P.T.: A review of mathematical modeling, artificial intelligence and datasets used in the study, prediction and management of Covid-19. *Appl. Intell.* **50**(11), 3913–3925 (2020)
 47. Nijsingh, N., van Bergen, A., Wild, V.: Applying a precautionary approach to mobile contact tracing for COVID-19: The value of reversibility. *Bioeth. Inquiry* **17**, 823–827 (2020). <https://doi.org/10.1007/s11673-020-10004-z>
 48. Noel, V., Ponce de León, M., Niarakis, A., Calzone, L., Valencia, A., Montagud, A.: PhysiBoss simulation of COVID19 infection (2020). <https://doi.org/10.21981/TQ16-VG65>. <https://nanohub.org/resources/pb4covid19>
 49. Noel, V., Leon, M.P.d., Niarakis, A., Calzone, L., Valencia, A., Montagud, A.: PhysiBoSS simulation of COVID19 infection (2020). <https://doi.org/10.21981/TQ16-VG65>. <https://nanohub.org/resources/pb4covid19>
 50. OECD: Trustworthy AI in Health (2020). <http://www.oecd.org/health/trustworthy-artificial-intelligence-in-health.pdf>
 51. Oliver, N., Lepri, B., Sterly, H., Lambiotte, R., Deletaille, S., De Nadai, M., Letouzé, E., Salah, A.A., Benjamins, R., Cattuto, C., et al.: Mobile phone data for informing public health actions across the covid-19 pandemic life cycle (2020)
 52. Pérez-Arnal, R., Conesa, D., Álvarez-Napagao, S., Suzumura, T., Català, M., Álvarez, E., Garcia-Gasulla, D.: Comparative analysis of geolocation information through mobile-devices under different Covid-19 mobility restriction patterns in Spain. *ISPRS Int. J. Geo Inf.* **10**(2), 73 (2021). <https://doi.org/10.3390/ijgi10020073>
 53. Phelan, J.C., Link, B.G., Tehranifar, P.: Social conditions as fundamental causes of health inequalities: theory, evidence, and policy implications. *J. Health Soc. Behav.* **51**(1_suppl), S28–S40 (2010)
 54. Pinter, G., Felde, I., Mosavi, A., Ghamisi, P., Gloaguen, R.: Covid-19 pandemic prediction for Hungary; a hybrid machine learning approach. *Mathematics* **8**(6), 890 (2020)
 55. Presidencia del Gobierno de España: Plan para la transición hacia una nueva normalidad (2020). <https://www.lamoncloa.gob.es/covid-19/Paginas/plan-transicion.aspx>
 56. Rodríguez, P., Graña, S., Alvarez-León, E.E., Battaglini, M., Darias, F.J., Hernán, M.A., López, R., Llana, P., Martín, M.C., Ramirez-Rubio, O., et al.: A population-based controlled experiment assessing the epidemiological impact of digital contact tracing. *Nat. Commun.* **12**(1), 1–6 (2021)
 57. Ruktanonchai, N., Ruktanonchai, C.W., Floyd, J., Tatem, A.: Using Google location history data to quantify fine-scale human mobility. *Int. J. Health Geogr.* **17**(28), (2018). <https://doi.org/10.1186/s12942-018-0150-z>
 58. Sambasivan, N., Kapania, S., Highfill, H., Akrong, D., Paritosh, P.K., Aroyo, L.M.: ‘Everyone wants to do the model work, not the data work: data cascades in high-stakes AI. In: SIGCHI, ACM (2021). <https://research.google/pubs/pub49953/>
 59. Santamaria, C., Sermi, F., Spyrtatos, S., Iacus, S.M., Annunziato, A., Tarchi, D., Vespe, M.: Measuring the impact of (covid-19) confinement measures on human mobility using mobile positioning data. A European regional analysis. *Saf. Sci.* **132**, (2020). <https://doi.org/10.1016/j.ssci.2020.104925>
 60. Savaget, P., Chiarini, T., Evans, S.: Empowering political participation through artificial intelligence. *Sci. Public Policy* **46**(3), 369–380 (2019)

61. SiĤa-Nowicka, K., Vandrol, J., Oshan, T., Long, J.A., Demřar, U., Fotheringham, A.S.: Analysis of human mobility patterns from GPS trajectories and contextual information. *Int. J. Geogr. Inf. Sci.* **30**(5), 881–906 (2016)
62. Sohrabi, C., Alsafi, Z., O’Neill, N., Khan, M., Kerwan, A., Al-Jabir, A., Iosifidis, C., Agha, R.: World health organization declares global emergency: a review of the 2019 novel coronavirus (Covid-19). *Int. J. Surg.* **76**, 71–76 (2020)
63. Suzumura, T., Kanezashi, H., Dholakia, M., Ishii, E., Napagao, S.A., P erez-Arnal, R., Garcia-Gasulla, D.: The impact of covid-19 on flight networks (2020). [arXiv:2006.02950](https://arxiv.org/abs/2006.02950)
64. Tangcharoensathien, V., Calleja, N., Nguyen, T., Purnat, T., D’Agostino, M., Garcia-Saiso, S., Landry, M., Rashidian, A., Hamilton, C., AbdAllah, A., Ghiga, I., Hill, A., Hougendobler, D., van Anandel, J., Nunn, M., Brooks, I., Sacco, P.L., De Domenico, M., Mai, P., Gruzd, A., Alaphilippe, A., Briand, S.: Framework for managing the COVID-19 Infodemic: methods and results of an online, Crowdsourced WHO Technical Consultation. *J. Med. Internet Res.* **22**(6), (2020). <https://doi.org/10.2196/19659>. <http://www.jmir.org/2020/6/e19659/>
65. Toch, E., Lerner, B., Ben-Zion, E., Ben-Gal, I.: Analyzing large-scale human mobility data: a survey of machine learning methods and applications. *Knowl. Inf. Syst.* **58**(3), 501–523 (2019)
66. Torres, A., Arguimbau, M., Bermejo-Mart n, J., Campo, R., Ceccato, A., Fernandez-Barat, L., Ferrer, R., Jarillo, N., Lorente-Balanza, J. ., Men ndez, R., et al.: Ciberesucicovid: a strategic project for a better understanding and clinical management of covid-19 in critical patients (2020)
67. Tzachor, A., Whittlestone, J., Sundaram, L.,  h Eigeartaigh, S.: Artificial Intelligence in a crisis needs ethics with urgency. *Nat. Mach. Intell.* **2**, 365–366 (2020). <https://doi.org/10.1038/s42256-020-0195-0>
68. UNICEF: Global Data Science Project for COVID-19 (2020). <https://www.covid19analytics.org/> (2020)
69. Vaishya, R., Javaid, M., Khan, I.H., Haleem, A.: Artificial intelligence (AI) applications for Covid-19 pandemic. *Diabetes & Metabolic Syndrome: Clin. Res. Rev.* **14**(4), 337–339 (2020). <https://doi.org/10.1016/j.dsx.2020.04.012>. URL <https://www.sciencedirect.com/science/article/pii/S1871402120300771>
70. Vannoni, M., McKee, M., Semenza, J., Bonell, C., Stuckler, D.: Using volunteered geographic information to assess mobility in the early phases of the (covid-19) pandemic: a cross-city time series analysis of 41 cities in 22 countries from march 2nd to 26th 2020. *Global Health* **16**, 85 (2020). <https://doi.org/10.1186/s12992-020-00598-9>
71. Wang, C., Ma, L., Li, R., Durrani, T.S., Zhang, H.: Exploring trajectory prediction through machine learning methods. *IEEE Access* **7**, 101441–101452 (2019)
72. Yang, Z., Zeng, Z., Wang, K., Wong, S.S., Liang, W., Zanin, M., Liu, P., Cao, X., Gao, Z., Mai, Z., et al.: Modified seir and ai prediction of the epidemics trend of covid-19 in china under public health interventions. *J. Thoracic Dis.* **12**(3), 165 (2020)