Technology in Society 71 (2022) 102137

Contents lists available at ScienceDirect

Technology in Society

journal homepage: www.elsevier.com/locate/techsoc

AI-powered public surveillance systems: why we (might) need them and how we want them

Catarina Fontes *, Ellen Hohma, Caitlin C. Corrigan, Christoph Lütge

Technical University of Munich, School of Social Sciences and Technology, Institute for Ethics in Artificial Intelligence, München, 80333, Germany

ARTICLE INFO

Keywords:
AI
Surveillance
Dataveillance
AI governance

ABSTRACT

In this article, we address the introduction of AI-powered surveillance systems in our society by looking at the deployment of real-time facial recognition technologies (FRT) in public spaces and public health surveillance technologies, in particular contact tracing applications. Both cases of surveillance technologies assist public authorities in the enforcement of the law by allowing the tracking of individual movements and extrapolating results towards monitoring and predicting social behavior. Therefore, they are considered as potentially useful tools in response to societal crises, such as those generated by crime and health related pandemics. To approach the assessment of the potentials and threats of such tools, we offer a framework with three dimensions. A function dimension, examines the type, quality and quantity of data the system needs to employ to work effectively.The consent dimension considers the user's right to be informed about and reject the use of surveillance, questioning whether consent is achievable and whether the user can decide fully autonomously/independently. Finally, a societal dimension that frames vulnerabilities and the impacts of the increased empowerment of established political regimes through new means to control populations based on data surveillance. Our analysis framework can assist public authorities in their decisions on how to design and deploy public surveillance tools in a way that enables compliance with the law while highlighting individual and societal tradeoffs.

1. Introduction

Considering the progressive datafication of reality, virtually everything can be coded and processed. With the possibility to collect, code, store and process data generated from activities that individuals carry out in everyday life, public authorities increase their fields of influence. Furthermore, new means enabled by AI support can assist complex decision-making processes and inform policies towards enhanced ways of managing and acting over public domains. All this encourages the uptake of AI systems leveraging big data.

As AI relies on the datafication process to progressively support humans in decision making processes in all spheres of life, personal data is used to generate intel. AI and specifically machine learning feed on available data to resolve patterns based on inference and validation. Data is the basis for the system to perform and deliver results, thus, we can state that data dependency is an inherent feature of AI. The nature of data and the generalization of its collection points expand the possibilities to monitor individual and societal activities/behavior on a large scale calling for a discussion around surveillance and its risks. Indeed, the opportunities for combining big data with automatization using AI are not without concerns.

Technologies that generate and transfer data are becoming embedded in our cities' landscapes or are already essential for conducting daily activities. Privacy is a major concern as everyone is subjected to ubiquitous surveillance, especially if we depart from the assumption that people generally wish to reveal as little about themselves as possible and want to control with whom they share certain information. Beyond potential threats to individual privacy, the abusive use or misuse of private data represents a risk to both individuals and communities as they come under the influence of third parties, be they public authorities or private enterprises. Consequently, individuals might see their ability to make free and uncoerced decisions hampered.

Even if efforts to implement regulatory frameworks were set in motion (see the GDPR (2018) and AI Act (2021) in the EU or in China the Data Security Law (DSL) and Personal Information Protection Law (PIPL) (2021)) the underlying principles remain under discussion and culturally dependent [1]. AI is still a moving target [2]. Therefore, it is crucial that we take on a role in monitoring and continuously

* Corresponding author. Arcisstr. 21, 80333 München, Germany.
E-mail address: catarina.fontes@tum.de (C. Fontes).

https://doi.org/10.1016/j.techsoc.2022.102137
Received 8 May 2022; Received in revised form 20 September 2022; Accepted 28 September 2022
Available online 1 October 2022

questioning whether the tradeoffs are acceptable.

In this paper, we question the reasoning behind AI-enabled public surveillance and outline an answer to why we actually might need automated surveillance systems and, if so, how we want them to behave and work. We describe two cases of AI-powered surveillance systems applied to healthcare and law enforcement: (1) contact tracing apps, particularly those adopted by many governments during the Covid-19 pandemic and (2) facial recognition technology (FRT) used to verify identities, remotely and in real time, widely tested and implemented around the world and highly contested or even banned in many cities and countries. Both cases enable the possibility to track individuals by collecting direct or indirect information on how they move through the world. Furthermore, they are tools that can potentially assist public authorities in law enforcement, leveraging what we describe in the next section as dataveillance.

From the examination of these two cases, we draw conclusions about potentials and threats of the use of AI-enabled public surveillance systems aimed at defining a three dimension framework. The function dimension examines the type, quality and quantity of data the system needs to work effectively. The consent dimension looks at the user's right to be informed on and reject surveillance, questioning whether consent is achievable and how it relates to individual autonomy. Finally, a societal dimension frames vulnerabilities and the impacts of the increased empowerment of established political regimes through new means to control populations based on data surveillance (or dataveillance).

2. The Use of Data for Public Surveillance

The process of transforming information generated by everyday life activities into quantified data has been referred to as datafication [3–5]. While not too long ago data was seen as a mere by-product or even something to be deleted as costs for storage surpassed its value, more recently it became its own currency [5] or capital [6]. With the ongoing digitalization, a rapidly increasing amount of data is available. The information is being 'datafied', which means the accumulation of unprecedented amounts of data, providing new possibilities to leverage information generated at both ends of the spectrum: personalization and generalization. Indeed, due to the imprinted knowledge on individuals, services and societies, it represents a highly valuable asset. Data translates to value, for instance, by enabling profiling and people tracking, optimizing systems, managing and controlling things, modeling probabilities and building and growing the value of assets [6].

Both governments and corporations have been increasingly enacting the use of data, crossing the line to the idea of mass surveillance of society [7]. According to Ref. [8], surveillance is a systematic form of attention with a purpose and implies both the idea of care and control. This definition is particularly interesting when we consider surveillance conducted by public authorities. There is a paternalistic sense behind this type of surveillance that goes along with the role of the state when caring for and protecting citizens through law and its enforcement. It also fits the idea of a broader agenda arguable in the definition of [9] that suggests dividing surveillance into direct surveillance, which aims at particular persons for particular reasons, and indirect surveillance, differing from the first by not having a specific target or purpose. Considering that the data can be stored and is therefore available for a longer period, a purpose can be defined and redefined at any point in time, so it is far from being a stable condition. Furthermore, a purpose is not necessarily what determines which data is to be generated and collected, once it can be a collateral result of using a specific service. Ref. [10] addresses the limits of surveillance in terms of how it curbs privacy and proposes a classification evolving from observation. The author considers passive observation, i.e. without intent to use the information to influence the target, a first type of surveillance. Thus, surveillance does not require an intentionality aspect, in terms of controlling others, albeit it encompasses leveraging the collected information in some way. Conversely, in active observation, collected information can be used retroactively namely to sanction the individual. However, to qualify as surveillance proper, the author implies that information must be used to prevent actions or to interfere in a proactive manner in an individual's behavior. Defining surveillance in the datafication era seems to twist the need for a predetermined and rather specific purpose, target or even starting point. However, it implies intentionality, considering that from a broader perspective leveraging data is what justifies surveillance in the first place.

Combining the long-lasting and ongoing trend of datafication with surveillance rationales results in the concept of dataveillance (see Table 1). Having data as the cornerstone, dataveillance is a type of surveillance enacted through sorting and sifting datasets in order to identify, monitor, track, regulate, predict and prescribe [11]. It differs from datafication due to the underlying observation intention. Both surveillance and dataveillance do not necessarily lead to inference but while the first implies a purpose and when done overtly affects people's behavior [10] dataveillance refers to continuous (and ubiquitous) tracking [5]. Moreover, dataveillance can be performed both on the personal and interpersonal level, can be voluntary or imposed and can be undertaken overtly or covertly. It implicates retrospective use of stored data, as well as data generated in real-time, due to interconnectedness, high speed processing capacities and automation applied to systems displayed in public and private spheres of an individual's life. In particular, covert dataveillance can bear high risks for societies, as these systems allow for little knowledge about how they operate and are often dispersed and constant in nature [12].

Table 1
Dataveillance definition.

	Datafication	Surveillance	Dataveillance	
Definition	Process of transforming reality features and social behavior into quantified data	Process of observing individuals or groups for a purpose and make inferences/ judgements on their behavior	Leveraging big data/ datafication to scale up/ramp up surveillance	
Attributes	Continuous and ubiquitous data generation/ collection using automatized systems	Retroactive, real-time and proactive observation of individuals or groups held overtly or covertly	Collecting and processing of personal data including in real-time by employing automatized systems and routines	
			Monitor, identify, track, regulate, predict, prescribe, prevent and steer individuals' or groups' (behavior)	Applications

Personal – but also non-personal – data comes with additional inherent knowledge about the person [13]. This poses two types of challenges: risks at an individual's level, which relate to personal privacy, and societal risks, involving the possibility of tracking each individual and making inferences about any person or group, having in profiling an emergent threat to individual autonomy and democratic values. From a technical perspective, a challenge lies in how data is interpreted during a sense-making process [14], as it is driven by mathematical interpretation that only secondarily considers the context [15] and biased outcomes can be difficult to identify and assess.

Many studies identify and define threats and risks of surveillance ramped up by big data and AI [5,7,10,12,13,16,17]. Our contribution in this field goes towards the outlining of a contingency strategy by presenting a first tool to assess threats and motivate stakeholders to question how a surveillance technology might hinder theirs/others individual liberties and choice. Firstly, to further discuss what purposeful surveillance means and when it crosses a red line to a mass, continuous and speculative kind, we present two AI-enabled systems adopted for health surveillance and public space surveillance. We examine how data is a condition to the functioning of those AI-powered public surveillance systems, the collection of which often blurs consent and transparency by deliberately or non-deliberately concealing risks. Secondly, we present three dimensions, pillars of a framework aimed at assisting public authorities in their decision to adopt tools that rely on dataveillance and raising awareness on individual and societal tradeoffs, liberating the space for an effective role for civil society.

2.1. From public health surveillance to health dataveillance

While in its beginnings public health surveillance activities were focused on the mere monitoring of diseases and reporting of resulting deaths [18], it has evolved into a "continuous, systematic collection, analysis and interpretation of health-related data" [19], or dataveillance. Epidemiologic observations have since developed from typically individual and locally concentrated health-related event recordings to large-scale, structured and preventive data interpretation [20]. This advancement has been facilitated by the availability and use of a growing variety of input opportunities.

While traditionally information is drawn from data sources like regularly repeated health surveys or disease registries, for example for monitoring cancer occurrences [20], new, innovative data sources are investigated and successively integrated in national health surveillance activities, e.g., syndromic data, like the number of patients that visit an emergency department [21]. Also, non-health related data, like social media, is increasingly regarded as a valuable information source. [22]; for example, suggested using Twitter geolocation data along with flight passenger information to identify potential illness risks, specifically, the spread of the Chikungunya virus in the Mediterranean area. Another example is the development of Google Flu Trends in 2008, a web application that predicted influenza occurrences based on linear regression between multiple, distinct Google search entries correlating with influenza time series data [23]. The inclusion of new input opportunities, especially involving Big Data, created a need to process these data using more innovative methods. This has led to the creation and growing adoption of AI-based systems supporting health dataveillance activities.

Text mining, for instance, is considered a powerful tool to extract disorganized information from electronic health records and, thus, enable an automated integration of diverse data types [24]. [25]; in this manner, proposed the use of Natural Language Processing (NLP) to automatically detect postoperative complications from clinical text documents. In particular, the Covid-19 pandemic has facilitated the development of innovative approaches of public health surveillance techniques, resulting in an increase in the use of smartphone apps for datafication and dataveillance purposes [4]. However, at the same time, it has revitalized the discussions on benefits and burdens of such mechanisms.

2.1.1. The case of contact tracing apps

With the spread of the Covid-19 pandemic, calls for innovative approaches to contain the virus spread quickly. Tracing apps were found to be one useful means to tackle the problem of reducing transmission of infectious diseases. Contact tracing, i.e., "identifying and monitoring each person who has been in contact with an infected person" [26]; p. 2), has already proven to be an effective measure in previous epidemics. Initially used for syphilis tracking in the 1930s, detecting potentially infected contacts of patients has been advanced during the Ebola outbreaks in West Africa in 2015 [27].

Traditionally, a human workforce is required, i.e., healthcare workers interview infected people, document and check provided contacts, as well as notify those at risk. However, this task is challenged by multiple limitations, such as low responsiveness of patients, memory gaps or bias of patients regarding their contact list, or the possibility to track encounters between persons unknown to each other [27]. To address these limitations and simultaneously support healthcare workers in their often standardized tasks, digital contact tracing applications have been developed to automatically and more effectively trace virus transmission.

Although there are multiple variations of the concrete design of tracing instruments, most smartphone-based contact tracing applications essentially follow the same principle. Wireless communication mechanisms are used to scan a user's surrounding for other transmitting smartphones and digital keys are exchanged to record the encounter of two devices. If a user submits a positive test result to the tracing app, all applications that store their key, and hence have been in contact with the infected person, receive a notification. Proximity and duration of the encounter determine the registered intensity of the contact and, therefore, may suggest to the notified user to self-isolate or seek medical advice.

Many implementation options have been proposed, varying, for example, on how data is collected, whether AI modules are used to calculate the infection risk or how the app can be installed on the users' smartphones [27]. Two major options for variation were profoundly discussed during the development phase of contact tracing apps: the approach to store user data and the method used to assess proximity.

Proposed strategies for data storage include centralized and decentralized techniques. While using centralized methods the user's contact list is uploaded to a platform and checks for potentially infected encounters are performed by a central entity, the decentralized methods retain user data on their smartphone and only perform certain pull and push requests to update the list of "positive-tested" user keys [28]. The implementation of the centralized version, while seemingly more straight-forward, however, bears potential for privacy issues as sensitive data leaves the user's disposition domain.

The way proximity is measured includes approaches such as WiFi MAC address sniffing, GPS, cellular network geolocation and Bluetooth tracing [28]. Bluetooth, with its advantage of measuring proximity based on signal strength, has been demonstrated to be highly effective. This is in particular the case because it does not allow the inference of geospatial location information [29], which is deemed to be more intrusive to users.

Regardless of their many variations in technical implementations, all contact tracing technologies offer strong potential in supporting the goal of registering exposures and informing contacts at risk during a pandemic. Due to automation, smartphone-based contact tracing apps are faster than their human counterparts. They can support human workforce and reduce the workload of healthcare authorities. Additionally, with increased speed they are able to administer higher numbers of infected people and contacts [30]. During the Covid-19 pandemic, authorities in Germany, for example, had to limit infection tracking to a minimum when incidences were too high for the tracing teams to handle [31]. Such problems can be prevented with the use of

automated tracking systems.

Another strong benefit of contact tracing technologies is their independence of user perception or knowledge. While human tracing teams are mostly limited to the patients' memory, technology-based tools can draw conclusions from previously recorded information. This improves the likelihood of registering risk cases, even facilitating the identification of people unknown to each other, an advantage that many governments sought to target during the Covid-19 pandemic. Ultimately, contact tracing technologies, hence, enable faster and better containment of communicable diseases leading to advancements for a society's health standards.

Many governments recognized these opportunities early on in the Covid-19 pandemic and relied on the development of national contact tracing applications to support their healthcare authorities. Singapore, for example, was one of the first countries to deploy a ready-to-use app to its citizens. 'TraceTogether' was published in March 2020 [29]. It uses short-distance Bluetooth signals to detect close contact with an infected user. Registration with a phone number is required, however, information on encounters is deleted after 21 days [32], which led to TraceTogether being broadly classified as well privacy preserving. A second example that was developed with a strong data privacy mindset was the German 'Corona-Warn-App'. High privacy debates during its development phase resulted in a decentralized data storage approach keeping most of the data on the user's phone instead of transmitting them to a governmental server. No personal identifier is needed to use the app and proximity is measured through location-independent Bluetooth signal recognition.

However, while many proposed tracing applications have put data privacy among the highest priorities, there are also examples where this intention was not properly realized. Aiming at a quick introduction of their app, Norway launched the first version of 'Smittestop' in April 2020, relying on live tracking of users' GPS coordinates and uploading them to a central server [33]. It was highly criticized for its substantial potential for mass surveillance, or dataveillance, as real-time transmission of user's geolocation data to a governmental database was found to be unbalanced with the required public health response [34]. Shortly after the privacy breach was detected, the Norwegian government decided to withdraw and revise their application.

Another example of how contact tracing technologies were implemented during the pandemic is the Chinese Alipay Health Code. Collected, along with self-reported data is analyzed to calculate a person's risk of being infected with Covid-19. Large debates arose with the launch of the application related to user autonomy, as adoption was mandatory, and a user's predicted risk determined their freedom of movement [32].

Despite its social and ethical risks, digital contact tracing, as a form of dataveillance, is generally found to be promising in helping contain virus spread [35–37]. Digital contact tracing enables, for instance, faster identification of users at risk, minimizing a testing and therefore isolation delay [35]. The various contact tracing applications developed during the Covid-19 pandemic show how technology can support in the response to such crises. Nevertheless, data protection is a concern and while some issues can be handled through regulation and control [38] society also has a claim in shaping such technological applications considering that effectiveness greatly depends on public acceptance.

Research has been pointing out factors that play a role in the public acceptance of more intrusive health surveillance measures, in particular looking at contact tracing technology. [39]; examine the extent to which public acceptance of contact tracing technology depends on cultural and socio-demographic aspects. The authors identified eight factors that influence the acceptance of such applications, including trust, privacy concerns or technological understanding [40]. investigate how different types of risk perception can either hinder or motivate an individuals' adherence to mobile health apps and focus on the risk-risk tradeoff with respect to privacy concerns or health-related hazards [17]. explore individuals' attitudes towards privacy and surveillance through a theoretical model to explain citizen' acceptance of governmental led surveillance and respective privacy protection measures. Besides their individual contributions, from a broader perspective, these findings further emphasize the interrelatedness between society and technology in a way that each is shaping the other. Therefore, developing guidance to deal with risks introduced by technically implementable concepts is paramount to ensure that benefits offset (at least most of) the burdens, justifying the deployment and fostering public acceptance.

2.2. The emergence of dataveillance of public spaces for law enforcement purposes

The datafication of everyday life activities in public spaces has been intended for multiple purposes, namely, to assist public authorities in managing urban issues. The adoption of a data-driven approach to urban governance is justified for being more objective and pragmatic in regard to making political decisions and guiding urban policy. Theoretically, data presents raw neutral facts, whereas ideologies create filters to interpret reality. Consequently, new mechanisms to collect and code public space features and activities in the form of urban big data have been implemented.

However, even when the amount justifies referring to it as big data, data is in fact partial, fragmentary and actually a product of complex socio-technical assemblages producing an incomplete and imperfect replication of reality [41,42]. This might point towards the idea that data is not neutral, but rather inherently biased, depending on the environment in which it was collected, the purposes behind its use and all the decisions about the way it is handled, processed, stored, analyzed and presented. Furthermore, socially embedded stereotypes that lead to biases must also be considered when employing big data. There are many forms of discrimination that serve as conditions or norms to organize societies and not necessarily considered unfair or negative. However, even in these cases, the scaling up and reproduction of biases due to the use of data based automatized systems may lead to unpredicted and undesirable outcomes.

Public space configures the spaces within a city that are accessible to everyone, independently of being publicly owned. This is the case for streets and squares, but also includes, for example, cinemas and sports halls. Moreover, they are characterized for being inclusive and fundamental in the creation/exaltation of cultural values and consummation of democracies. Public spaces in cities are also places of conflict where power is contested and social relations are established [43,44]. Public space is, therefore, a place of public sovereignty, managed and mediated by public authorities in charge of maintaining public order and enforcing the law. Surveillance is a mechanism to assist public authorities in doing that, namely by supporting the policing of streets. The deployment of cameras (CCTV) across cities, for instance, was considered a big step towards leveraging technology to the surveillance of public spaces.

Overall surveillance is also changing the nature of public space because it affects the way people act [45,46]. Through the use of CCTV systems, public authorities acquired a tool to support policing and law enforcement activities. Nevertheless, in the last decades, many studies have pointed out that these systems might not be as efficient as hoped and that specific circumstances should be accounted for prior to deciding whether CCTV is the best solution to foster safety and reduce crime in public spaces [46–48]. Additionally, ethical issues arise when employing CCTV related infringements to individual privacy and also transparency, discrimination and exclusion [49–52].

2.2.1. The case of facial recognition technology (FRT)

Facial recognition technology (FRT) enables identity validation by measuring and analyzing a person's facial image and comparing it against other samples in a database. This technology is quickly bridging the gap between traditional surveillance and dataveillance in public spaces. The processing delivers a score indicating the likelihood of the

compared images referring to the same person. The system can be used for authentication, which means verification of identity, identification or establishment of identity and tracking a face on several recorded images. Additionally, the process can take place "in real time" or on pre-recorded materials. The applications for FRT are multiple. Identifying faces on photos on social networks, validating identity working as a password or to verify attendances, diagnosing diseases and evaluating candidates in recruiting are just a few examples [53–55].

These technologies also have the potential to enhance CCTV systems by allowing real time identity checks. This means that AI is used to analyze the data and make certain inferences and decisions in real time, enabling new possibilities for the dataveillance of public spaces in support of law enforcement and public security.

Public authorities have already begun justifying the use of Smart CCTV systems (i.e., upgraded CCTV systems due to the use of AI-enabled FRT) to potentially assist in policing and reducing criminality by identifying known or suspected criminals, terrorists or missing persons or for tracking down suspicious behavior. The use of such a system implicates the continuous monitoring of public spaces and consequently surveillance of human activities and presences, raising pressing concerns on privacy and potential threats to individual rights and liberties (FRA, 2019; [56,57].

Despite the technology's regulation being at a preliminary stage (the European Commission has only recently proposed its regulation through the AI Act [58]), the use of FRT for public surveillance has been rapidly spreading, with sixty-four countries testing and adopting it in some way [59,60]. At the other extreme many countries and cities have been claiming and declaring the ban of the use of FRT for public surveillance. Exemplifying the contentious nature of this form of dataveillance.

In key examples, the Leicestershire Police in the UK used automated facial recognition in 2015 at the Download Festival, checking 90,000 people against a Europol watchlist. The South Wales Police has been using FRT to monitor big events such as outdoor festivals, sports events or public protests [61]. According to public reports, between 2017 and 2019, over 70 events were targeted, and 60 persons were arrested. The Metropolitan Police Service has already used FRT to monitor public spaces and conducted 10 test deployments between 2016 and 2019 to verify technical accuracy and to assess implications for policing operations [62].

In Germany in 2016, the Cologne Police deployed 26 stationary video cameras at the main station forecourt and other central public areas, such as the surroundings of the Cologne Cathedral and Breslauer Platz. The Hamburg police trialed the use of FRT during the G20 Summit in 2018. In 2017, FRT was tested at Berlin Südkreuz train station and about three years later the federal government and Deutsche Bahn announced the deployment of cameras with possible recourse to real-time FRT as part of increased security measures at train stations [63].

In Serbia and Kenya, the Huawei Safe City program employed FRT as well. In Belgrade, the plan was to deploy facial recognition and license plate recognition cameras in 800 locations across the city. This project faced strong opposition from individuals and organizations in civil society. Nairobi deployed 1800 HD cameras and 200 HD traffic surveillance systems as part of the same program [60]. The AI Global Surveillance Index [64] covers even more examples from Canada to China and from Mexico to India, looking at seventy-five countries listed for employing AI surveillance. Of those, sixty-four are already actively incorporating facial recognition systems in their AI surveillance programs. Thus, the employment of FRT for dataveillance is not an abstract idea, but an increasingly common occurrence. On the other hand, more and more cities have put bans or moratoriums on governmental uses of facial recognition. The trend is particularly visible in the US [65–67]. While in the US *cities* are leading the response to handling the assessment of threats versus benefits of such systems, in Europe responses are being prepared at the *international* level, with the EU developing regulation to standardize norms and protect the fundamental rights and European values.

The European Commission has recently proposed an AI Act in which the use of real time remote biometric identification systems in publicly accessible spaces is highly conditioned. Implementation is only accepted in exceptionable situations, such as searching for victims of crime, prevention of a substantial and imminent threat to life or physical safety of natural persons or of a terrorist attack, detection, localization, identification or prosecution of a perpetrator or suspect of a criminal punishable in the Member State concerned by a custodial sentence or a detention order for a maximum period of at least three years [58]. The European Data Protection Board (EDPB) and the European Data Protection Supervisor (EDPS) go even further on prohibiting the use of such systems in public spaces calling for a general ban on any use of AI for an automated recognition of human features in publicly accessible spaces, such as faces, gait, fingerprints, DNA, voice, keystrokes and other biometric or behavioral signals, in any context. They argue that remote biometric identification of individuals in publicly accessible spaces poses a high-risk of intrusion into individuals' private lives, with severe effects on the populations' expectation of being anonymous in public spaces [68]; pp.2-3).

Even if public authorities seem to have plausible arguments to justify the mass surveillance, or dataveillance, of public spaces in support of law enforcement, and according to some studies people might be willing to accept being exposed to FRT for this purpose [69,70], its deployment raises questions in terms of fairness, transparency and proportionality [52,71]. Indeed, counter-arguments suggest that its deployment may be beyond the rule of law [72,73], pose a threat to fundamental and human rights [56,57], lead to the undermining of democratic values due to a potential chilling effect [62,74] and can impactthe social meanings of public spaces by promoting the exclusion of vulnerable groups [16,50,75]. Hence, actually assessing the need for its use in a systematic way is a relevant task for governments and policymakers.

3. Assessing the ethics and public dataveillance: balancing benefits and threats

Dataveillance technologies have inherent benefits for helping societies manage crises or preventing crimes. However, as has been illustrated by the examples in the previous section, the adoption of these technologies also bears a cost. In order to understand what is at stake, we have been pointing out how data, particularly its personal/private nature, is a drawback when the goal is to design and deploy technologies that respect human rights and more ideally are people-centered. The latter assumption means that a reason to adopt a specific technology is beyond considering it a technical solution to a specific problem, even if deemed harmless to people. The adoption of a specific technology should contend to a societal demand regardless of who is the end-user and ensure from a holistic perspective that it is meant for the "good" of humankind.

In order to mitigate risks around the use of personal data, privacy requirements and impacted users' consent are preliminary conditions and anonymization is key not to overstepping privacy standards. Additionally, even if a technology is for some reason made inactive or removed (including malfunctioning), this does not mean that the data and information it provided are likewise out of use nor nullified. Independently of how it was generated, data can be reused by whoever has ever had access to it, heightening the need for adequate security measures. In Fig. 1, we exemplify what a data cycle can look like.

Data and an AI system lifecycles do not follow the same timeline. Data becomes an independent entity, i.e. it can be copied, multiplied, manipulated and repurposed. The [76] defined the AI system lifecycle in four phases: design, data and models; verification and validation; deployment; operation and monitoring. The first phase includes planning and design, data collection and processing, as well as model building. Therefore, during the development of an AI system there can be an immediate need to use available data for training that was likely collected by another system for a different purpose.

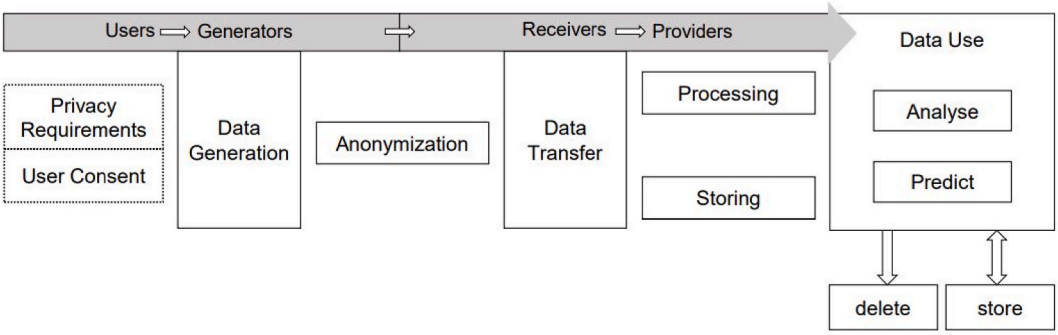

Fig. 1. Example of a data cycle.

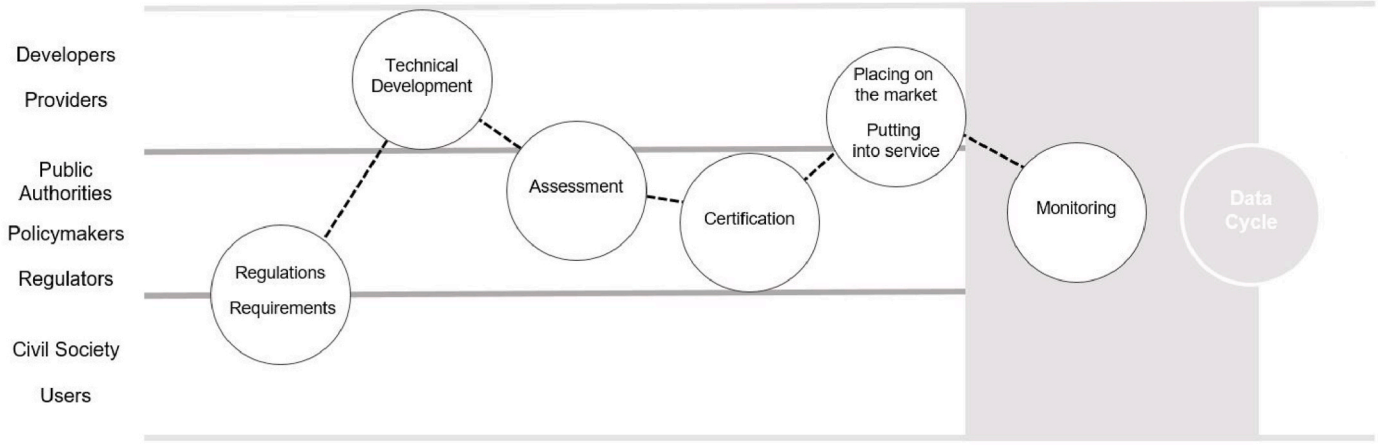

Fig. 2. Towards an AI governance framework. Identifying stakeholders and grading responsibility in an AI system lifecycle.

In its current version, the AI Act [58] refers to the AI lifecycle defined through what we could call milestones. In Fig. 2, we propose setting them in an abstract timeline while weighing stakeholders' responsibility. We further illustrate how an AI system lifecycle and a data cycle encompass different steps and even if they intersect, they do not completely overlap.

3.1. The functional dimension: assessing the data privacy/fairness trade-off

Data dependency is an inherent feature of AI, as AI-enabled tools require it to function. This means not only that the system depends on the availability of data to perform, but also relies on it to deliver. Thus, data is paramount in terms of amount and quality.

In terms of quantity, the question is how much data is enough? The kind of data needed to support the functioning of many surveillance technologies (including the two presented - FRT and contact tracing apps) is of personal nature, i.e., data is produced by individuals while pursuing their everyday activities and contains features that have the potential to trace the data back to the individual itself. Collecting more data heightens the risk of progressively intruding the individuals' privacy. This raises privacy issues and supports the argument of only collecting the minimum amount of data needed for functioning.

Privacy is a societal value whose loss represents a pitfall. It relates to individuals' access to, and control over, how personal data is used [77]. Collecting personal data implies pushing the borders of privacy so that an invasive data collecting process triggers the necessity of moving towards increasingly intrusive surveillance and potentially leading to infringements to privacy as an acquired right.

The case of contact tracing apps illustrates the premise. For instance, the German contact tracing tool Corona-Warn-App was designed with the highest data privacy requirements, aiming to inhibit any inference of movement patterns through the Smartphone application [78,79]. Furthermore, it was meant to conceal the location where exposures might have occurred in order to avoid discrimination against and exclusion of infected people. The steps taken to protect the individuals' privacy by not collecting and not enabling such information present, however, limits the tool's efficiency by impairing and complicating the contact tracing work of health authorities. Namely, in the event of exposure signaled by the app, it would be impossible to determine whether the risk of being infected actually exists or if the conditions of an encounter might reduce or neutralize risks (e.g., if it took place with further hygiene precautions) [78]. This caused some decisions made by public authorities to be not based on accurate information, but rather based on incomplete and inconclusive data.

Privacy protection and the emphasis on overcoming the idea of tracing an individual's movements under false pretexts led to certain generalizations that might inadvertently result in enforcing unnecessary, and thus unfair, restrictive measures hampering individual liberties. Should the collection of more data have been consented, the overall efficiency of the tool would have been improved towards potentially fairer results and more accurate predictions that could further help with the response to a pandemic.

The more data a system uses the better prepared it will be to deliver high quality results. Consequently, larger amounts of data will positively impact accuracy and overall efficiency, increasing fairness in how users are perceived and treated. However, in order to avoid malfunctioning or erroneous outputs, using not only more data, but also highly representative data is a basic condition. Therefore, it is not exclusively a question of amount, but cumulatively a question of data quality.

Enough representative data would ideally generate an identical "digital twin" reality that enables precise simulations and, again, more accurate and efficient outputs. However, deriving a perfect digital copy of the reality's inherent processes can be troublesome.

Inadequate data collection approaches can be the source of an insufficient representation of particular groups [80–82]. In the presented cases, this can have many facets. Considering that smartphone-based contact tracing applications require the necessary technical devices, a considerable target group, e.g., elderly people, cannot be reached and hence are not well represented in the analyzed datasets [83]. Acknowledging that the worldwide mobile services subscription is 66% of the world's population [84], this is not only an issue of excluding specific groups.

As for FRT, the place of deployment can per se lead to biased results by over-targeting certain groups within society. Indeed, as the tool is considered a potential solution for crime, it is more likely that public authorities can justify its deployment in areas more affected by or susceptible to criminality. This in turn would lead to even higher arrest rates in areas employing the technology, as it would have increased observation, further stigmatizing the neighborhood and disproportionately affecting those residents [85].

Moreover, even when/if this identical replica of a reality is achieved, it does not necessarily work towards fairness. Biased data can be caused by lack of representativeness, but it is not only a data collection problem. While aiming at a replication of our realities, we tolerate incorporating the reproduction of embedded cultural biases, which often take place at the grassroots' level when it comes to unfair kinds of discrimination.

Deploying a system by which the face of every passer-by is analyzed, mapped and their identity verified is a decision that impacts the whole population, meaning that everyone using public space is under surveillance without being a suspect of any crime. Theoretically, everyone in a space under surveillance would then be exposed to the same risks that ultimately result in an AI-powered FRT triggering police action. But, the fact is that the likelihood of a false positive (false match) or false negative (failed association) has been shown to depend on ethnicity and gender [86,87].

According to the EU Ethics Guidelines [88] fairness means ensuring equal and just distribution of both benefits and costs, and that individuals and groups are free from unfair bias, discrimination and stigmatization. The guidelines also consider that if unfair biases can be avoided, AI-enabled systems could even increase societal fairness. In practice, due to lack of data quality that may be rooted on socially embedded patterns of discrimination, certain groups within society are more likely to be flagged by the system, justifying being stopped and asked to prove identity. This means being at a higher risk of unfair discrimination and potentially overstepping of the presumption of innocence, thus the denial of their rights.

The principles of data privacy and fairness that underlie the dimension of functionality, require data processing measures that, in fact, pull in opposite directions. While trying to balance out both privacy and fairness, privacy might be the malleable notion, as tradeoffs may justify understanding it in light of levels of trust in the technology and public authorities and acknowledged needs [89].

3.2. The consent dimension: a paradox

Transparency refers to the extent of information made available for those that were not part of a process in order to grant them the possibility to make informed decisions [90]. It relates to the idea of making knowledge accessible by creating relationships of trust between different stakeholders. When applied to the implementation of AI-enabled systems, it means, for instance, that the users are informed about what the system does, and how and why it does it. Therefore, transparency is important to build confidence in the technology [91].

Transparency and consent come hand in hand when discussing the relationship between individuals and governments, namely when it comes to the collection and use of personal data. However, the complexity of the terms and of the technologies discussed here create hurdles to transparency and highly affect the meaning of consent. Consent goes beyond awareness. It implies an action of agreeing and simultaneously creating the path to withdraw. It has, therefore, a dynamic nature.

In the use of personal data, consent should be both conditional and dynamic. This means that the processing of a person's data must be explicitly called out, allowing them to cancel a service and potentially rescind any data shared [91]. If we extrapolate to privacy as a right applied to the case of FRT, as mentioned, people might be able to compromise and accept more intrusive surveillance if they can feel safer in public space in return. It is the same situation in the case of contact tracing apps. If they are proved to help in containing the spread of a virus, it becomes a sort of social duty to put the protection of lives above individual privacy.

The latter argument illustrates another paradox of consent, on the one hand it should be dynamic, i.e., there must be a way to withdraw, and, on the other hand, it is conditional but should not be conditioned. For instance, if the values at stake are presented as unquestionable, then the possibility of a diverse opinion is conditioned and the autonomy to make an uncoerced choice is hindered. When a technology that relies on the intrusion of individual privacy is presented by public authorities as the most suitable solution to societal problems, such as criminality and terrorism or the spread of a disease, the request for consent is blurred by the apparent lack of arguments to decline. Notwithstanding, even if a person gives consent to share personal data with public authorities for a specific end, the lack of capacity to understand the technology and how a system will leverage the data is a further challenge for transparency.

In another example from the public health domain, the German Infection Protection Act (IfSG) specifies in § 6 that measles, chickenpox and recently also Covid-19 must be reported to the national health agency if identified by a doctor or laboratory. To protect the broader society, although seemingly personal, patients are not asked for consent to report such information. With the introduction of new technologies, questioning to what extent automating identification and reporting of diseases is acceptable becomes a matter of consent over surveillance. However, the complexity of the situation presents nuances if we consider the sources of information available. Non-traditional data sources such as social media/networks (e.g., Twitter) detain valuable personal data that can support public authorities in a vast number of ways. They enable the association of identity and location, if the user makes these available and increase the representation of certain population groups towards an improved comprehensive tracking of reportable diseases. Nevertheless, it is questionable if this would be a transparent use of such information, even if made publicly available, as a Twitter user was not consulted and, especially, has not consented to make the data available for health surveillance purposes.

Whereas contact tracing apps require an active action from individuals to install the application on their private smartphones, which can be considered a kind of consent if use is not mandatory, the deployment of FRT in public spaces raises other issues in terms of how to discuss consent. Firstly, the deployment might not be that obvious as the end-interface that people will eventually recognize are cameras, sometimes discretely added to the urban landscape. Often FRT is implemented as an upgrade to CCTV so not an evident or even noticeable change for citizens crossing public space.

On the one hand, full disclosure in policing activities can be counterproductive, affecting the effectiveness and pertinence of surveillance in a sort of paradoxical situation where transparency and efficiency seem impossible to balance out towards justifying the use of such technology [52]. On the other hand, if people exposed to such systems are not aware that they are under surveillance, the need to consent to it appears redundant. Hence, unless public authorities increment transparency by disclosing the location and extent surveillance is being conducted, open a discussion to reasoning the purposes and obtain

consent, citizens become either unaware of its existence or start distrusting the technology, which is already motivating community backlash in several countries (see for example the European campaign Reclaim Your Face).

Additionally, refusing to consent to being under surveillance in public space is possibly equivalent to giving up on access to it. This contravenes the inclusivity of public space, potentially promoting the exclusion of groups particularly exposed and vulnerable to surveillance [75]. Furthermore, individual decisions are conditioned by the uncertainty of being able to refuse or withdraw consent without penalty [62], so if denying being under surveillance is understood as connected to being excluded from accessing to public spaces, it may be seen as a dissonance to pragmatically understand the meaning of consent. Without access to adequate information, consent is redundant since it can only be granted with enough autonomy where a choice can be truly uncoerced. This also relates to the idea that alternatives were considered and are acceptable, which means they do not deprive individuals from benefiting the plenitude of their privileges and rights such as access to goods and services.

3.3. The societal dimension: dataveillance and the risk of abusive use

As explained above, the lack of transparency and underestimating the need for consent already constitutes an abusive use of personal data by disrespecting individual privacy and freedom of choice regardless of the type of tools in which the data is employed. However, the new possibilities enabled by AI bestow an unprecedented amount of power on the watcher [74]. This has already been recognized in existing AI ethics approaches, such as the IEEE's Principles for Ethically Aligned Design, in which creators are advised to "guard against all potential misuses and risks of A/IS (autonomous and intelligent systems) in operation" [91]; p. 31).

The risks triggered by augmenting power imbalances, which dataveillance tools have the potential to enable go beyond immediate infringements of individual privacy. The fact that data is stored leads to risks which initial consent does not efface. Due to a breach or lack of security, information may be accessed by people that were not initially intended to have access and consequently, reused in ways that were not initially anticipated [10].

Further advancements have been suggested as being possible to deliver increasingly complex interpretations and predictions towards emotion analysis and drawing scenarios for future actions. AI would be able to analyze biological and behavioral patterns and make inferences on what is "normal" or "abnormal" [57,92–95].

At a societal level, mass and automated surveillance may radically affect the way societies function and have the potential to undermine democracy. FRT, for example, makes it possible for a government to conduct real-time location tracking and behavior policing of an entire population. However, the applications go even beyond identification or identity verification, as the processing of biometric data can be used for profiling individuals by collecting and categorizing personal characteristics such as age, sex and ethnic origin. These profiles can inform comprehensive behavioral analyses, the basis for the generation of reputational scores for social credit systems [96]. The repurposing of stored data, has been trialed for example in Chinese social credit system projects [97] leveraging existing public information collected for other purposes.

Concerning public health dataveillance, the Covid-19 pandemic has proven that governments are willing to undertake drastic measures in order to protect populations from the spread of diseases. Protecting populations might controversially mean depriving individuals from certain acquired rights, such as the freedom of movement and discrimination of citizens according to their adherence to governmental norms, as the case of policies based around vaccination status have illustrated.

Another example is the "Kwarantanna Domowa" ("Home Quarantine") App that the Polish Ministry of Digitalization has implemented to alleviate and reduce costs of home visits by the police and national health authorities for self-isolation compliance checks [98]. The initially voluntary, but later mandatory, tool uses geolocation tracking and facial recognition technology to validate the user's current location and identity in order to remotely monitor whether an imposed quarantine is actually being adhered to. To ensure correct functioning users must upload a selfie taken at their self-isolation place upon app installation, which will be used as a reference photo for identity verification in the course of their quarantine. The app then randomly asks the user from time to time during the self-isolation period to take and send selfies within an allotted time period of 20 min from their quarantine place. If the photos are not transmitted in time, the app is designed to automatically inform the responsible authorities.

The two investigated examples might portray situations of power abuse prompted by AI-powered technologies for public surveillance. Abuse in this case is due to public authorities' claim to have the authority to control over image and location data of citizens, i.e., disregarding the citizens' personal data privacy and autonomy by not transparently informing them about the collection and use of personal information and, without allowing them the choice to self-decide whether this is acceptable or not. This also shows that individual consent might not always offset abusive use, since prerequisites such as individual autonomy and transparency, are overlooked.

In a scenario where public dataveillance tends towards omnipresence, the empowerment of public authorities over citizens by accessing extensive information on individual and societal activities creates a background to enforce and coerce populations to accept the restraining of rights and further strengthen authoritarianism and autocracy. The banalization of surveillance by naturalizing the use of intrusive surveillance systems may develop a chilling effect in a progressive path to transitioning individuals' emancipation and social empowerment towards coerced societies with increasingly constrained rights and autonomy.

To prevent the banalization of the adoption of intrusive dataveillance systems and abusive use of personal data, there is the need to raise the awareness of public authorities and populations about benefits and threats of the use of these systems, while empowering both to be able to discuss tradeoffs. This can be done through top-down initiatives such as frameworks and regulation (e.g. EU Ethics Guidelines, GDPR), but in parallel through bottom-up movements, namely led by civil society organizations in a position to represent public opinion and comprehend impacts for societies. Civil society has indeed a paramount role to play in steering, demanding and enforcing democratic practices together with ethical safeguards prior to employing dataveillance related mechanisms.

Conducting trials, studying impacts and collecting feedback through independent reports [62] has been a way to move forward in the use of FRT to monitor public spaces in the UK. Simultaneously civil society organizations (such as Big Brother Watch and LIberty), media and research institutions (as [69] have been adding to the discussion by raising public awareness and measuring the pulse of public opinion. The proposed EU Artificial Intelligence Act (Regulation 2021/0106) bans the use of "'real-time' remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement", only opening room for strictly necessary purposes as for the search for victims of crime, the prevention of a threat to life or physical safety or of terrorist attacks, or the search for perpetrator or suspect of a criminal offence (Art. 5(1) (d), AI Act). This sort of containment response reveals the per se concerns that this kind of AI-powered tool represents to established democratic systems and looks at the mitigation of undesirable impacts by assessing risks and inhibiting negative repercussions.

In the case of contact tracing apps, the freedom to use or not to use these applications in the scope of Covid-19 has been highly debated since the beginning of the pandemic, with the result that most countries promoted a voluntary installation [33]. The German Federal Ministry of Health accepted a delay in setting up the official national contact tracing

application as a tool to control the spread of Covid-19 in order to check proof data privacy requirements and security standards. To avoid "mission creep", many countries have explicitly limited the use to its intended purpose and for a particular timeframe. However, ongoing reconsideration and questioning of the design and implementation choices is inevitable to detect and mend potential omissions. The case of Norway's national contact tracing tool shows that governmental power abuse is not just a matter of intention but that unintended privacy breaches can - and have to be - spotted, communicated and repaired. The example furthermore underlines the value and strength of good governance and democracy by continuously scrutinizing governmental measures.

To summarize, the increasing use of mass and automated surveillance systems on populations bestows unprecedented power on public authorities and governments by enabling access to privileged personal data of citizens that might be further used for multiple purposes and combined with other records. The given examples demonstrate that governments and supranational organizations have been acknowledging risks and setting in motion containment mechanisms to ensure the emphasis is put on the use of "AI for good" and giving individuals the assurance that their concerns are being tackled and their rights are safeguarded.

3.4. A framework for dataveillance analysis

The ODCE's framework to classify AI systems [99] presents both the advantages and downsides of a holistic approach. There are four dimensions (people and planet, economic context, data and input, AI model, task and output) each subdivided in multiple sub-dimensions that relate to a specific question. The dimension data and input, the closest dimension to our focus, brings about the following questions: Are the data and input collected by humans, automated sensors or both? Are the data and input from experts; provided, observed, synthetic or derived? Are the data dynamic, static, dynamic updated from time to time or real-time? Are the data proprietary, public or personal data (related to identifiable individual)? If personal data, are they anonymized; pseudonymised? Are the data structured, semi-structured, complex structured or unstructured? Is the format of the data and metadata standardised or non-standardised? What is the dataset's scale? Is the dataset fit for purpose? Is the sample size adequate? Is it representative and complete enough? How noisy are the data?

We argue that by defining a stakeholder's perspective, there is a shift in the way we can formulate those questions above, where the answers are not redundant, but rather lead to a more complete view of the underlying risk. On another note, the dimensions we propose relate to the presented use cases driven by a gradual generalization towards an outline of the framework. As a background to laying down a framework for dataveillance analysis, we define the following key takeaways based on the three dimensions examined and presented above:

- There are limits to what is acceptable within democratic societies in terms of using personal data and conducting surveillance activities, regardless of the purpose.
- There needs to be an openness to debate. Balancing out pros and cons is a process that not only includes, but focuses equally on the civil society perspective.
- Establishing data privacy, transparency and individual autonomy are priority requirements to implement dataveillance systems.
- Continuous reevaluation of the system is vital. This means questioning whether the system is still the best solution, ensuring that surveillance is still necessary and responding to the initially intended purposes.

Given these findings, Table 2 outlines some core questions that both policymakers considering the use of dataveillance systems and civil society actors potentially impacted by such systems should ask themselves.

Table 2
– Questions for policymakers and civil society. Core questions that policymakers and the civil society should consider regarding the use of dataveillance systems according to the three proposed dimensions.

	Policymakers' perspective	Civil Society's perspective
The function dimension	How much data is needed to ensure accuracy while minimizing intrusion on privacy?	Is personal data collected? Which kind of data and to what extent?
	Can we ensure data anonymization?	Is data timely anonymized and secured?
	Are we looking at personalization or generalization as a data processing outcome?	Is bias and discrimination against individuals or groups tackled and mitigated?
The consent dimension	Has consent to use data been clearly established?	Have we been timely and appropriately informed on how, why and what personal data is collected and asked for consent?
	Is this informed consent dynamic and continuously available?	Have we given consent?
	Is this informed consent conditional, if so, how may this affect autonomy?	Are we aware of the possibility of consent withdrawal? Are getaways clear and effective?
The societal dimension	What are the potentials for misuse of the system?	Are we being targeted beyond acceptable limits or overexposed to surveillance?
	What mechanisms can be put in place to avoid this misuse?	Are the purposes behind the system's implementation clear?
	Has civil society been informed and consulted during the process?	Are individual rights safeguarded in the proposal or trade-offs explained?
		Do we trust the technology? Do we trust public authorities?

The presented cases (i.e. real-time facial recognition technologies and contact tracing apps) showcase the use of surveillance technologies to assist public authorities in the enforcement of the law by allowing tracking individuals' movements and extrapolating results towards monitoring and predicting social behavior. Therefore, they are useful tools in response to societal crises such as those generated by crime and health-related pandemics. While these systems represent advancements and bring up new possibilities to the way public authorities deal with threats to human lives, they simultaneously pose new challenges to existing balance and order by calling for a compromise on individual privacy and even autonomy. By asking the questions above before employing of a system that relies on dataveillance, government policymakers and civil society will be in a position to systematically consider three relevant dimensions of the overarching questions of do we need automated surveillance systems and, if so, how do we want them to behave and work. This proceess can help to maximize the usefulness of these systems for bettering society and managing crises, while minimizing their negative impact on democratic values and cultural norms.

4. Conclusion

In this article, we address the introduction of AI-powered surveillance systems, or dataveillance systems, in our society by looking at the deployment of real-time facial recognition technologies in public spaces and contact tracing apps used for public health management. While the surveillance of populations might be necessary for law enforcement purposes, AI-enabled surveillance systems may trigger the offset of established order and societal values. This does not mean that benefits will not outstrip potential threats nor that populations are not willing to accept the tradeoffs, but rather that public authorities become ever more responsible to self-regulate their own power, as well as the power asserted by tech enterprises for providing the tools to collect and handle personal data generated through everyday life activities.

The question is not only how much power can be handed to AI

systems, but also how can we ensure that the power given to public authorities over individuals will be used exclusively in their best interest. Adopting AI-powered surveillance systems expose populations to an increased risk of power imbalance, based on the enabling of access to privileged information on individuals' private lives collected within and feeding the systems to provide intel to public authorities. While the output information may be extremely relevant to predict, prevent and neutralize societal threats, it pushes the boundaries of personal privacy and bestows an immense amount of power on public authorities over individuals, which may lead to the undermining of democratic values and individual rights and liberties.

Therefore, it is crucial to understand how personal data is being handled, by whom and for what purposes, so that specific actors become accountable to regulate and supervise the overstepping of acceptable uses and the consented limits for surveillance. Transparency is key and so is fairness. Even if the whole population is under surveillance, specific groups are still disproportionately affected and exposed due to the way they use or interact in public space. Individuals and groups in civil society have, therefore, a role to play by engaging in the debate over the technology's impacts and limits and scrutinizing their governments.

To move forward in assessing the context in which the AI-powered public surveillance systems are employed, and understand if the benefits outweigh the costs in each particular situation, a multi-stakeholder and multi-dimensional framework would need to be build. However, as the OECD framework [99] showcases, it is rather challenging starting from scratch and building a simultaneously generic and detailed enough tool. Thus, our contribution in this paper may help to inform further initiatives to develop frameworks both from a methodological point of view and as a starting point to future work.

Declarations of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgements

Funding: This work was supported by the Institute for Ethics in Artificial Intelligence (IEAI) at the Technical University of Munich.

References

[1] Y.-H. Kao, S.G. Sapp, The effect of cultural values and institutional trust on public perceptions of government use of network surveillance, Technol. Soc. 70 (C) (2022), https://doi.org/10.1016/j.techsoc.2022.102047.
[2] C. Bartneck, C. Lütge, A. Wagner, S. Welsh, What is AI?, in: An Introduction to Ethics in Robotics and AI. SpringerBriefs in Ethics Springer, 2020.
[3] K. Cukier, V. Mayer-Schoenberger, The rise of big data: how it's changing the way we think about the world, The Best Writing on Mathematics 2014 (2013) 20–32.
[4] C.S. Lee, Contact tracing apps for self-quarantine in South Korea: rethinking datafication and dataveillance in the COVID-19 age, Online Information Review 45 (4) (2021) 810–829.
[5] J. van Dijck, Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology, Surveill. Soc. 12 (2) (2014) 197–208.
[6] J. Sadowski, When data is capital: datafication, accumulation, and extraction, Big data & society 6 (1) (2019), 2053951718820549.
[7] R. Martínez-Béjar, G. Brändle, Contemporary technology management practices for facilitating social regulation and surveillance, Technol. Soc. 54 (2018) 139–148, https://doi.org/10.1016/j.techsoc.2018.04.003.
[8] D. Lyon, Surveillance Society. Monitoring Everyday Life, Open University Press, Buckingham and Philadelphia, 2001.
[9] A.F. Westin, Privacy and freedom, Wash. Lee Law Rev. 25 (1) (1968).
[10] H.S. Sætra, Freedom under the gaze of Big Brother: preparing the grounds for a liberal defence of privacy in the era of Big Data, Technol. Soc. 58 (2019) 101160.
[11] R. Clarke, Information technology and dataveillance, Commun. ACM 31 (5) (1988) 498–512.
[12] D. Lupton, M. Michael, 'Depends on who's got the data': public understandings of personal digital dataveillance, Surveill. Soc. 15 (2) (2017) 254–268.
[13] J.-E. Mai, Big data privacy: the datafication of personal information, Inf. Soc. 32 (3) (2016) 192–199.
[14] M. Lycett, 'Datafication': making sense of (big) data in a complex world, Eur. J. Inf. Syst. 22 (2013) 381–386.
[15] C. Anderson, The end of theory: the data deluge makes the scientific method obsolete, Wired magazine (2008).
[16] M. Hirose, Privacy in public spaces: the reasonable expectation of privacy against the dragnet use of facial recognition technology, Conn. Law Rev. (2017) 377.
[17] A. Ioannou, I. Tussyadiah, Privacy and surveillance attitudes during health crises: acceptance of surveillance and privacy protection behaviours, Technol. Soc. 67 (2021), 101774.
[18] S.M. Teutsch, R.E. Churchill, Principles and Practice of Public Health Surveillance, Oxford University Press, USA, 2000.
[19] World Health Organization, Surveillance in Emergencies, World Health Organization, 2021. https://www.who.int/emergencies/surveillance. (Accessed 28 April 2022).
[20] L.M. Lee, S.B. Thacker, Public health surveillance and knowing about health in the context of growing sources of health data, Am. J. Prev. Med. 41 (6) (2011) 636–640.
[21] A. Chiolero, D. Buckeridge, Glossary for public health surveillance in the age of data public health surveillance in the age of data science, J. Epidemiol. Community Health 74 (7) (2020) 612–616.
[22] J. Rocklöv, Y. Tozan, A. Ramadona, M.O. Sewe, B. Sudre, J. Garrido, C.B. de Saint Lary, W. Lohr, J.C. Semenza, Using big data to monitor the introduction and spread of Chikungunya, Europe, 2017, Emerg. Infect. Dis. 25 (6) (2019) 1041–1049, https://doi.org/10.3201/eid2506.180138.
[23] A.E. Aiello, A. Renson, P.N. Zivich, Social media–and internet-based disease media–and internet-based disease surveillance for public health, Annual Review of Surveillance for Public Health. Annual Review of Public Health 41 (2020) 101–118.
[24] Q. Bi, K.E. Goodman, J. Kaminsky, J. Lessler, Q. Bi, K.E. Goodman, J. Kaminsky, J. Lessler, What is machine learning? A primer for the epidemiologist, American Journal of Epidemiology 188 (12) (2019) 2222–2239.
[25] H.J. Murff, F. FitzHenry, M.E. Matheny, N. Gentry, K.L. Kotter, K. Crimin, R. S. Dittus, A.K. Rosen, P.L. Elkin, S.H. Brown, Automated identification of postoperative complications within an electronic medical record using natural language processing, JAMA 306 (8) (2011) 848–855.
[26] C. Perscheid, J. Benzler, C. Hermann, M. Janke, D. Moyer, T. Laedtke, O. Adeoye, K. Denecke, G. Kirchner, S. Beermann, N. Schwarz, D. Tom-Aba, G. Krause, Ebola outbreak containment: real-time task and resource coordination with SORMAS, Frontiers in ICT 5 (7) (2018) 1–11, https://doi.org/10.3389/fict.2018.00007.
[27] S. Jacob, J. Lawarée, The adoption of contact tracing applications of COVID-19 by European governments, Policy Design and Practice 4 (1) (2021) 44–58.
[28] S. McLachlan, P. Lucas, K. Dube, G.S. McLachlan, G.A. Hitman, M. Osman, N. Fenton, The Fundamental Limitations of COVID-19 Contact Tracing Methods and How to Resolve Them with a Bayesian Network Approach. London, UK, 2020, 27042.66243.
[29] N. Ahmed, R.A. Michelin, W. Xue, S. Ruj, R. Malaney, S.S. Kanhere, A. Seneviratne, W. Hu, H. Janicke, S.K. Jha, A survey of COVID-19 contact tracing apps, IEEE Access 8 (2020) 134577–134601.
[30] U. Lee, A. Kim, Benefits of mobile contact tracing on COVID-19: tracing capacity perspectives, Front. Public Health 9 (586615) (2021), https://doi.org/10.3389/fpubh.2021.586615.
[31] Berlin.de, Health Authorities Reduce Contact Tracing, Berlin.de, 2022 https://www.berlin.de/en/news/coronavirus/7257547-6098215-health-departments-reduce-contact-tracin.en.html. (Accessed 28 April 2022).
[32] R. Jalabneh, H.Z. Syed, S. Pillai, E.H. Apu, M.R. Hussein, R. Kabir, S.M.Y. Arafat, M.A.A. Majumder, S.K. Saxena, Use of mobile phone apps for contact tracing to control the COVID-19 pandemic: a literature review, in: S. Nandan Mohanty, S. K. Saxena, S. Satpathy, J.M. Chatterjee (Eds.), Applications of Artificial Intelligence in COVID-19, Springer Singapore, 2021, pp. 389–404, https://doi.org/10.1007/978-981-15-7317-0_19.
[33] A. Boch, C. Corrigan, Ethics and the use of AI-based tracing tools to manage the COVID-19 pandemic, TUM IEAI Research Brief, 2020. June 2020, https://ieai.mcts.tum.de/wp-content/uploads/2020/06/Research-Brief_ContactTracingAppsFinal-1.pdf.
[34] Amnesty, Bahrain, Kuwait and Norway contact tracing apps among most dangerous for privacy. https://www.amnesty.org/en/latest/news/2020/06/bahrai

n-kuwait-norway-contact-tracing-apps-danger-for-privacy/, 2020. (Accessed 28 April 2022).
[35] M.E. Kretzschmar, G. Rozhnova, M.C. Bootsma, M. van Boven, J.H. van de Wijgert, M.J. Bonten, Impact of delays on effectiveness of contact tracing strategies for COVID-19: a modelling study, Lancet Public Health 5 (8) (2020) 452–459.
[36] M. Salathé, C.L. Althaus, N. Anderegg, D. Antonioli, T. Ballouz, E. Bugnion, S. Capkun, D. Jackson, S.-I. Kim, J.R. Larus, Early evidence of effectiveness of digital contact tracing for SARS-CoV-2 in Switzerland, Swiss Med. Wkly. 150 (2020).
[37] V. Shubina, A. Ometov, A. Basiri, E.S. Lohan, Effectiveness modelling of digital contact-tracing solutions for tackling the COVID-19 pandemic, J. Navig. 74 (4) (2021) 853–886.
[38] EDPB, Guidelines 04/2020 on the use of location data and contact tracing tools in the context of the COVID-19 outbreak. https://edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_20200420_contact_tracing_covid_with_annex_en.pdf, 2020. (Accessed 4 September 2022).
[39] M. Villius Zetterholm, Y. Lin, P. Jokela, Digital contact tracing applications during COVID-19: a scoping review about public acceptance, in: Informatics, 8, MDPI, 2021, 48.
[40] C.D. Tran, T.T. Nguyen, Health vs. privacy? The risk-risk tradeoff in using COVID-19 contact-tracing apps, Technol. Soc. 67 (2021), 101755.
[41] R. Kitchin, The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences, SAGE, London, 2014.
[42] R. Kitchin, Data-driven Urbanism, in: R. Kitchin, T.P. Lauriault, G. McArdle (Eds.), Data and the City, Routledge, London and New York, 2018, pp. 44–56.
[43] H. Lefebvre. The Production of Space, Blackwell, Oxford UK and Cambridge USA, 1991.
[44] D. Mitchell, The end of public space: people's park, definitions of the public and democracy, Ann. Assoc. Am. Geogr. 85 (1995) 108–133.
[45] M. Foucault, Discipline and Punish: the Birth of a Prison, London House, New York, 1979.
[46] H. Koskela, The gaze without eyes: video-surveillance and the changing nature of urban space. Progress in, Hum. Geogr. 24 (2) (2000) 243–265.
[47] N.R. Fyfe, J. Bannister, The eyes upon the street: closed-circuit television surveillance and the city, in: N.R. FYFE (Ed.), Images of the Street: Representation, Experience and Control in Public Space, Routledge, London, 1998, pp. 254–267.
[48] J. Ditton, E. Short, Yes, it works, no, it doesn't: comparing the effects of open street CCTV in two adjacent Scottish Town Centres, in: Surveillance, Crime and Social Control, Routledge, 2006, pp. 201–223.
[49] N. Taylor, State surveillance and the right to privacy, Surveill. Soc. 1 (1) (2002) 66–85.
[50] H. Koskela, Cam Era– the contemporary urban panopticon, Surveill. Soc. 1 (3) (2003) 292–313.
[51] C. Norris, G. Armstrong, CCTV and the social structuring of surveillance, in: C. Norris, D. Surveillance Wilson (Eds.), Crime and Social Control, Routledge, 2006, pp. 157–178.
[52] C. Fontes, C. Perrone, Ethics of Surveillance: Harnessing the Use of Live Facial Recognition Technologies in Public Spaces for Law Enforcement, TUM IEAI Research Brief, 2021. December 2021, https://ieai.mcts.tum.de/wp-content/uploads/2021/12/ResearchBrief_December_Fontes-1.pdf.
[53] B. Jeon, B. Jeong, S. Jee, Y. Huang, Y. Kim, G. Park, T. Choi, A facial recognition mobile app for patient safety and biometric identification: design, development, and validation, JMIR Mhealth Uhealth 7 (4) (2019), e11472, https://doi.org/10.2196/11472.
[54] S. Sawhney, K. Kacker, S. Jain, S.N. Singh, R. Garg, Real-time Smart attendance system using face recognition techniques, in: 9th international conference on cloud computing, data science & engineering (Confluence), IEEE, 2019.
[55] Y.S. Su, H.Y. Suen, K.E. Hung, Predicting behavioral competencies automatically from facial expressions in real-time video-recorded interviews, J Real-Time Image Proc 18 (2021) 1011–1021, https://doi.org/10.1007/s11554-021-01071-5.
[56] London Policing Ethics Panel (LPEP), Final Report on Live Facial Recognition, LPEP, 2019. London, http://www.policingethicspanel.london/reports.html. (Accessed 28 April 2022).
[57] E. Pauwels, Artificial Intelligence and Data Capture Technologies in Violence and Conflict Prevention Opportunities and Challenges for the International Community, Global Center on Cooperative Security, 2020.
[58] European Commission (COM), Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, 2021. Brussels.
[59] P. Brey, Ethical aspects of facial recognition systems in public places, Info, Comm & Ethics in Society 2 (2004) 97–109.
[60] S. Feldstein, The Global Expansion of AI Surveillance, Carnegie Endowment for International Peace, 2019.
[61] B. Davies, M. Innes, A. Dawson, Universities' Police Science Institute Crime & Security Research Institute, Cardiff University, 2018 (last accessed April 28, 2022), https://afr.south-wales.police.uk/wp-content/uploads/2019/10/AFR-EVALUATION-REPORT-FINAL-SEPTEMBER-2018.pdf.
[62] P. Fussey, D. Murray, Independent Report on the London Metropolitan Police Service's Trial of Live Facial Recognition Technology, University of Essex, Human Rights Centre, 2019. Available at: https://48ba3m4eh2bf2sksp43rq8kk-wpengine.netdna-ssl.com/wp-content/uploads/2019/07/London-Met-Police-Trial-of-Facial-Recognition-Tech-Report.pdf (last accessed April 28, 2022).
[63] L. Montag, R. Mcleod, L. De Mets, M. Gauld, F. Rodger, M. Pełka, The rise and rise of biometric mass surveillance in the EU. EDRi (European digital rights) and EIJI (edinburgh international justice initiative), Brussels (2021).
[64] S. Feldstein, AI Global Surveillance Index, Carnegie Endowment for International Peace, 2019.
[65] K. Conger, R. Fausset, S.F. Kovaleski, San Francisco Bans Facial Recognition Technology, The New York Times, May 14, 2019.
[66] S. Ravani, Oakland Bans Use of Facial Recognition Technology, Citing Bias Concerns, San Francisco Chronicle, July 16, 2019.
[67] K. Lannan, Somerville Bans Government Use of Facial Recognition Tech, WBUR, 2019. June 28,.
[68] EDPB, EDPS, Joint Opinion 5/2021 on the proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) (2021). Available at: https://edpb.europa.eu/system/files/2021-06/edpb-edps_joint_opinion_ai_regulation_en.pdf. (Accessed 28 April 2022).
[69] Ada Lovelace Institute, Beyond face value: public attitudes to facial recognition technology, 2019. Report and survey data, https://www.adalovelaceinstitute.org/report/beyond-face-value-public-attitudes-to-facial-recognition-technology/. (Accessed 28 April 2022).
[70] G. Kostka, L. Steinacker, M. Meckel, Publ. Understand. Sci. 30 (6) (2021) 671–690, https://doi.org/10.1177/09636625211001555.
[71] D. Castelvecchi, Is facial recognition too biased to be let loose? The technology is improving — but the bigger issue is how it's used, Nature 587 (2020) 347–349, https://doi.org/10.1038/d41586-020-03186-4.
[72] Big Brother Watch, Face off. The Lawless Growth of Facial Recognition in UK Policing, 2018. https://bigbrotherwatch.org.uk/wp-content/uploads/2018/05/Face-Off-final-digital-1.pdf. (Accessed 28 April 2022).
[73] R. van Brakel, How to Watch the watchers? Democratic oversight of algorithmic police surveillance in Belgium, Surveill. Soc. 19 (2) (2021) 228–240.
[74] E. Selinger, W. Hartzog, The inconsentability of facial surveillance, Loyola Law Rev. 66 (2019) 101–122.
[75] C. Fontes, C. Lütge, Surveillance and power relations. The use of facial recognition technologies and remote biometric identification in public spaces and impacts on public life, Direito Público 18 (100) (2022). https://www.portaldeperiodicos.idp.edu.br/direitopublico/article/view/6203.
[76] OECD, Scoping the OECD AI principles: deliberations of the expert group on artificial intelligence at the OECD (AIGO), in: OECD Digital Economy Papers, 291, OECD Publishing, Paris, 2019.
[77] L. Floridi, J. Cowls, M. Beltrametti, R. Chatila, P. Chazerand, V. Dignum, C. Lütge, E. Vayena, AI4People –an ethical framework for a good society: opportunities, risks, principles, and recommendations, Minds Mach. 28 (2018) 689–707.
[78] M. Laaff, Corona-warn-app – app trifft amt, Zeit Online (2020). https://www.zeit.de/digital/2020-07/corona-warn-app-gesundheitsamt-hotline-labor-anbindung/komplettansicht. (Accessed 28 April 2022).
[79] Robert Koch-Institut, Infektionsketten digital unterbrechen mit der Corona-Warn-App – Die Corona-Warn-App ist ein wichtiger Baustein der Pandemiebekämpfung, Robert Koch- Institut, 2021. https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/WarnApp/Warn_App.html. (Accessed 28 April 2022).
[80] U. Gasser, M. Ienca, J. Scheibner, J. Sleigh, E. Vayena, Digital tools against COVID-19: taxonomy, ethical challenges, and navigation aid, The Lancet Digital Health 2 (8) (2020) 425–434.
[81] C. Klingler, D.S. Silva, C. Schuermann, A.A. Reis, A. Saxena, D. Strech, Ethical issues in public health surveillance: a systematic qualitative review, BMC Publ. Health 17 (1) (2017) 295.
[82] L. Lucaj, C. Corrigan, Ethical implications of the use of AI to manage the COVID-19 outbreak, TUM IEAI Research Brief (2020). April 2020, https://ieai.mcts.tum.de/wp-content/uploads/2020/04/April-2020-IEAI-Research-Brief_Covid-19-FINAL.pdf.
[83] E. Hohma, Assessing Fairness in AI-Enabled Public Health Surveillance, TUM IEAI Research Brief, 2021. October 2021, https://ieai.mcts.tum.de/wp-content/uploads/2021/10/ResearchBrief_October2021_Assessing-Fairness-in-AI-enabled-Health-Surveillance-FINAL-V2.pdf.
[84] GSMA, The Mobile Economy 2021, GSMA, 2021. https://www.gsma.com/mobileeconomy/wp-content/uploads/2021/07/GSMA_MobileEconomy2021_3.pdf. (Accessed 28 April 2022).
[85] K. Lum, W. Isaac, To predict and serve? Significance 13 (5) (2016) 14–19.
[86] J. Buolamwini, T. Gebru, Gender shades: intersectional accuracy disparities in commercial gender classification, Proceedings of Machine Learning Research 81 (2018) 1–15.

[87] P. Grother, M. Ngan, K. Hanaoka, Face Recognition Vendor Test Part 3: Demographic Effects, U.S. Department of Commerce - National Institute of Standards and Technology, 2019.
[88] European Commission (COM), Ethics guidelines for trustworthy AI. https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai, 2019. (Accessed 28 April 2022).
[89] J. Hong, J. Landay, An architecture for privacy-sensitive ubiquitous computing, in: Proc. 2nd Int'l Conf. Mobile Systems, Applications, and Services, ACM Press, 2004, pp. 177–189.
[90] A.M. Florini, The battle over transparency, in: A.M. Florini (Ed.), The Right to Know: Transparency for an Open World, Columbia University Press, New York, 2007, pp. 1–16.
[91] The IEEE Initiative on Ethics of Autonomous and Intelligent Systems (IEEE), Ethically Aligned Design, IEEE, 2017. https://ethicsinaction.ieee.org/wp-content/uploads/ead1e.pdf. (Accessed 28 April 2022).
[92] M.S. Ryoo, Human Activity Prediction: Early Recognition of Ongoing Activities from Streaming Videos, IEEE International Conference on Computer Vision, Barcelona, Spain, 2011.
[93] Y. Kong, Y.R. Fu, Human Action Recognition and Prediction: A Survey, 2018. *ArXiv,* abs/1806.11230.
[94] D. Yang, A. Alsadoon, P.W.C. Prasad, A.K. Singh, A. Elchouemi, An emotion recognition model based on facial recognition in virtual learning environment, Procedia Comput. Sci. 125 (2018) 2–10, https://doi.org/10.1016/j.procs.2017.12.003.
[95] T. Singh, D.K. Vishwakarma, Video benchmarks of human action datasets: a review, Artif. Intell. Rev. 52 (2019) 1107–1154, https://doi.org/10.1007/s10462-018-9651-1.
[96] P. Langer, Lessons from China - the formation of a social credit system: profiling, reputation scoring, social engineering, in: The 21st Annual International Conference on Digital Government Research, 2020, pp. 164–174, https://doi.org/10.1145/3396956.3396962.
[97] X. Dai, Toward a Reputation State: the Social Credit System Project of China, 2018. Available at: SSRN 3193577.
[98] EDRi, The rise and rise of biometric mass surveillance in the EU. https://edri.org/wp-content/uploads/2021/07/EDRI_RISE_REPORT.pdf, 2021. (Accessed 28 April 2022).
[99] OECD, OECD framework for the classification of AI systems, in: OECD Digital Economy Papers, vol. 323, OECD Publishing, Paris, 2022, https://doi.org/10.1787/eedfee77-en.