DOI: 10.53116/pgaflr.7108

Government Cybersurveillance and AI: A New Equation

Michael M. Losavio*

* Associate Professor, Department of Criminal Justice and Department of Computer Science and Engineering, University of Louisville, Louisville, Kentucky, USA, e-mail: mmlosa01@louisville.edu

Submitted: 30 November, 2023 | Accepted: 2 April, 2024 | Published online: 24 April 2024

Abstract: There is a tension between state oversight and state intrusion into our personal lives. The analytical powers of artificial intelligence/machine learning and the pervasive data collection of the Internet of Things, the Smart City and all those personal devices we use together permit revelations as to our lives as never seen before. We must consider the impact of this on the relations between citizen and government in this new, ubiquitous world of government cybersurveillance and revelation.

Keywords: surveillance, cybersurveillance, artificial intelligence, algorithms

1. Technology and privacy

There is comfort, as children, in being watched over by family and friends as we play, do homework, sleep. That sometimes changes as we grow older. For some being surveilled by a parent, or the state, becomes a gross intrusion into our privacy and personal autonomy. The concept itself comes from the French "sur", meaning "over" and "veiller", meaning "to watch". For others, being watched over remains a blanket of safety in a world of evolving threats. Those relationships between citizen and state now evolve in a data world of constant surveillance. Information for that may come from data generated by the Internet of Things to Smart City administrative systems, from locational data of a cellphone to networks of closed-circuit television cameras. The huge, usually voluntary engagement with social media creates vast troves of information on peoples' lives. Much of that is open to public view, and easily accessible through a law enforcement request or search warrant where there is probable cause for a crime. It may also lead to concern and litigation of the impact of such systems on mental well-being. The San Mateo Board of Education is suing YouTube and other social media systems for

causing emotional and psychological injury to its minor students.[1] The impact of such massive surveillance may affect those relationships across all domains of society. Some see it as the future of policing and governmental public safety (Davidson, 2019). Others see it as a means to "predict enemies" (Deeks, 2018).

Jones v. United States[2] presented the United States Supreme Court with the question of the legality and propriety of inexpensive and easy computer mediated tracking via small devices placed on a suspect's automobile by law enforcement. The Court found such tracking invaded the privacy of the person tracked and would require a court order/search warrant upon a showing of probable cause of a crime, but decided the matter on traditional trespass grounds. Anticipating issues to come with government computer-mediated surveillance, Associate Justice Sonya Sotomayor opined in her concurring opinion that such extensive surveillance technology may very well come to change the relations between citizen and state.

The police powers of the state are as fundamental as they are intrusive, involving investigation, detention and punishment of people. As in many other areas of human activity, the role of computing – data collection and analysis – in policing has been growing. In particular, the use of systems for "predictive policing" use data to forecast areas of risk and, in some matters, individuals of risk who may be subject to special scrutiny and police attention. That additional scrutiny may become a self-fulfilling prophecy of guilt, even where there is none.

Computing has changed many aspects of our lives, especially our privacy, autonomy and safety. It changes the power of the government to see into the lives of its citizens in unprecedented ways. This has led to jurisprudence seeking to protect people in the face of new technology, just as where the question of privacy into the ruling in *Kyllo v. United States*[3] barring infrared surveillance of a home without a warrant, even where there was no physical intrusion to the home. The line of cases from *Jones v. United States, Riley v. California*[4] and *Carpenter v. United States*[5] have reinforced privacy rights against our use of new computing applications, and their use by law enforcement to generate unprecedented information on the lives of others in the exercise of the police power of the state. Legislative protections lag in the United States, especially when compared to those of the European Union and its General Data Protection Regulation (hereinafter: GDPR).[6]

There are similar concerns as to commercial surveillance and data security practices that may harm consumers through intrusions into their private lives.[7] Regulations may look at how commercial organizations may (1) collect, aggregate, protect, use, analyse and retain consumer data; and (2) transfer, share, sell, or otherwise monetise that data in ways

[1] *San Mateo County Board of Education and Nancy Magee, in her official capacity as San Mateo County Superintendent of Schools v. YouTube, LLC, Google LC, XXVI Holding Inc., Alphabet Inc., Snap Inc., TikTok Inc. and Bytedance Inc.*, United States District Court for the Northern District of California, Case No 3:23-cv-01108, pp. 1–107.

[2] *Jones v. United States,* 565 U.S. 400 (2012).

[3] *Kyllo v. United States,* 533 U.S. 27, 34 (2001).

[4] *Riley v. California,* 573 U.S. 373, 381 (2014).

[5] *Carpenter v. United States,* 138 S. Ct. 2206, 2221 (2018).

[6] Regulation (EU) 2016/679 (General Data Protection Regulation), OJ L 119, 04.05.2016; cor. OJ L 127, 23.5.2018.

[7] U.S. Fed. Trade Comm'n, Commercial Surveillance and Data Security Rulemaking, 16 CFR Part 464 Trade Regulation Rule on Commercial Surveillance and Data Security Rulemaking (Aug. 11, 2022).

that are unfair or deceptive; in an extreme example of private-state data collaboration the U.S. military purchased the entire satellite surveillance imagery of an active battlefield by a private company, preventing others from accessing it during the battle (Campbell, 2001).

In turn, private surveillance methods implicate ways in which government surveillance may use those private surveillance resources to look into the lives of people (Campbell, 2001). That may yet implicate the privacy principles of *Katz v. United States* that the Fourth Amendment, the primary statutory protection of citizens' privacy in the United States, protects *people,* not places.[8] The principles in *Katz* led to the rule in *Riley* that personal electronic device could not be searched absent a warrant, and then to *Carpenter* that historical cell site location information on a cellphone user's activity, though collected and held by a private third party, could not be accessed as to track that user absent a properly issued search warrant. While U.S. statutes protect communications in real-time and stored electronic communications for a period of time, there are few other protections from government surveillance in the new data age from federal legislation. Yet these can be instructive. Chapter 121 of the U.S. Criminal Code, 18 USC §2701, et seq., addresses stored wire and electronic communications and transactional records access. It outlines a framework for guide government access to such data and procedures to do so as to offer privacy protections to the people involved with such communications.

The potential data space that supports surveillance keeps growing, with increases in the sources of data and means to store it for access and use. These are paralleled by increasing growth in analytical power to pluck that data from "practical obscurity" and put it to surveillance use. Those analytic systems have grown in power – artificial intelligence, machine learning, neural networks – to the point that they can infer from that data aspects of and conclusions about peoples' lives that, in the past, would have been difficult or impracticable to obtain. These analytical systems for public security are used around the world, and their use is increasing.

One survey of police in the United States found that 88% of respondents used data tools for police purposes. (International Association of Chiefs of Police, 2011). Such data collection, transmission and analysis can engage and support police resources from the investigation of singular criminal activity to ways in which law enforcement engages with its community (Losavio & Losavio, 2014). These systems can promote quick, timely identification and apprehension of suspects, as well as mis-identify the innocent as offenders (CBS, 2013; Bensinger & Chang, 2013; Dyer, 2013; Connors & Zauderer, 2013). Recognition of these issues have led, in some cases, to law enforcement policies and training on the proper acquisition and use of such data, including open source social media and local sensing networks (Stuart, 2013).[9] It seems that fewer and fewer human activities can escape police oversight once a person steps out the door of their home, physically or virtually. The 2013 Rand Corporation study *Predictive Policing* examined the use of data analytics in public safety across the United States (Perry et al., 2013). Its observations showed a broad application of data analytics to a variety of public safety/crime issues.

[8] *Katz v. United States*, 389 U.S. 347 (1967).

[9] United States Department of Justice Global Justice Information Sharing Initiative Federal Advisory Committee, 2010. Online: https://shorturl.at/jwFNR

Some of these systems of police analytics were effective. Others, however, were less so, raising issues as to whether or not they should even be used.

With these new computing technologies, we are presented with five fact domains that may impact the intrusiveness of government surveillance. And how those intrusions may be mitigated if we wish or need to do so.

Five technical concerns should be considered in relation to new computing technologies and their use in government surveillance; those are:

1. What are the data at issue, and its data subject?
2. How and where are those data sensed, perceived and generated?
3. How, where and under what conditions are those data collected and stored?
4. How are such data accessed and analysed?
5. How are those data and analysis used?

These domains reflect constitutional, statutory and common law privacy issues under U.S. law, and similar characterisations under the General Data Privacy Regulation of the European Union and the proposed EU Artificial Intelligence Law.

Table 1 details how a technical node may risk compromise as to particular privacy and security interests:

Table 1.
Risk nodes mapped to privacy and security

Risk node	**Privacy breach**	**Security breach**
Sensors	Intrusion	Hijacked/spoofed instrumentation, or erroneous data
Networks	Transfer beyond control	Interception, masquerade, hostile injection
Analytics	Revelation, mis-inference	Mis-inference, false negatives, false positives

Source: Compiled by the author.

Such compromise may be matched to outcomes from that compromise as to direct where protections for personal privacy and security could be directed for the most effective protection of those interests from attack.

These reflect points at which regulation may be used to protect privacy and security from new areas of surveillance. Regulation as to conduct and the engineering of these systems may offer privacy protection in this computational age through limitations on the use of these technologies in adverse ways. A regulatory control may mandate technical and engineering protections, such as the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule[10] on protection obligations, its Security Rule[11] on required controls for electronic personal health information and the GDPR restrictions on transfer of personal data across borders. Technical regulation can support privacy and personal autonomy, though the engineering must be flexible to permit regulatory audit and

[10] HIPAA Standards for Privacy of Individually Identifiable Health Information, 45 CFR Parts 160 and 164 (2000) (US). Online: https://t.ly/BJplX

[11] HIPAA Security Rule, 45 CFR Part 160 and Subparts A and C of Part 164 (US) (2010).

response; the HIPAA Security Rule requires entities under the jurisdiction of the Security Rule, such as health care providers and those they contract with that access personal health information (PHI) must:

- ensure the confidentiality, integrity, and availability of all e-PHI they create, receive, maintain or transmit
- identify and protect against reasonably anticipated threats to the security or integrity of the information
- protect against reasonably anticipated, impermissible uses or disclosures
- ensure compliance by their workforce

The ways in which persons may be injured by the improper use of such information, including PHI, as to need regulation for protection is seen in Table 2 below. Table 2 details governance of technology and system use as to provide controls that, in turn, may protect against particular injuries to personal privacy and autonomy.

Table 2.
Governance of technical operations and use creating protections

Systems issues	**Regulatory controls**	**Privacy injuries mitigated**
Sensed and networked data →	Use limitations and filters, controls over transmission →	Improper publicity
Analytics against the networked and collected data →	Data limits, bi-directional network limits, use limits →	Intrusion into private affairs
Incomplete data, network gaps and flawed analytics →	Assurance criteria for data, vetted algorithms →	Mis-inference, false assertions, false light

Source: Losavio et al., 2018

Whether by the government or private parties, new systems of data-driven surveillance can change the dimensions of the space of personal autonomy, shrinking it more and more. It may indeed change the relationships between people, civil society and government where people no longer feel secure from their government nor from malicious private parties.

2. The data at issue, its data subject and the provenance of data for surveillance

Data about individuals are the primary issue and concern in state surveillance as it provides direct evidence of activities and inferences and conclusions about a person. These may range from where they were at a particular time or what they may have purchased that could be put to malicious purpose. This data can be compiled to create a profile on

a person as to impact the individual as a data subject and as to their person and professional life. That data may directly connect to a data subject or permit inferences about that data subject, to varying degrees of specificity and identification. Closed circuit video or an investigator's images of a person may connect directly to a specific person, correctly or in error. Credit card use, payment systems, cellphone use and automobile license plate readers provide a one-off inference that possessor/owner of the related technology is using it. All these data are subject to analysis by analytical engines of computation.

These data objects, depending on system design, may be accompanied by related metadata such as time, place, activities and associations of the data subject. As the amount of information expands, more direct and inferential conclusions may be made about an individual. Whether those conclusions are correct or not represent the particular risk in their use. This may lead to potentially significant damages to the person that is connected to those conclusions. A significant risk now comes from the reliability of artificial intelligence systems in assembling and analysing data; this risk may be heightened by the lack of transparency and the ability of those who build such systems to explain the operations of those AI systems. These factors can impact the measurement and detection of such risks as well as creating them.

Data are the starting point, the grist for the analytical mill. Artificial intelligence brings a renewed relevance to the old computational saying "Garbage In, Garbage Out". The provenance and quality of the data used to develop AI systems and used by them to produce results are vital concerns. They are critical for weighing the reliability of a system in making any conclusions at all, as required by judicial rules of evidence. In the world of semi-supervised and unsupervised machine learning, AI systems teach themselves through their analysis of large bodies of data, such as Large Language Models. Bad and corrupt data, and data based on improper, illegal or unfair rules can lead to the creation of badly corrupted algorithms. Those corrupted algorithms may then generate corrupt and erroneous conclusions about their data subjects. And the targets of a police investigation.

Even where AI systems are built on good data with algorithms that perform as programmed, problems may develop in how those systems are used absent ways to assure the reliability of the data output of the systems. This is vitally important and a growing challenge with the rise of "deep fake" technologies that make forensic determinations of authenticity more and more difficulty. Images, audio and video may all be subject to "deep fake" manipulation and fabrication, even as images, audio and video represent some of the most powerful evidence considered by a finder of fact such as the jury. Because of this new challenge to authenticity, it becomes vital to see and judge how data is created and collected, and possible solutions to the problems of provenance in a time of deep fakes.

3. How data is sensed, perceived and generated for surveillance

A starting point for validating provenance and reliability, if possible, is how data is generated and created. Where data is created by an electronic system there must be some evidence of the reliability of that evidence. The accuracy and reliability of the sensing or

perceiving system that generates that data is of primary importance. As that data may not be directly authenticated by a human observer who can testify that the data and representations are accurate and representative, other evidence of reliability will be needed.

System testing and evaluation such as required for the validation of expert evidence under rules like the Federal Rules Evidence 702 and 703 of the United States may establish the reliability of that data sensing and generation. For electronic and digital systems this may be particularly important as the complexity of the systems increases. For artificial intelligence systems this is particularly critical due to issues with the transparency and "explainability" as to how such systems operate and produce results and conclusions. These issues may be difficult to resolve as to assure reliability under rules for judicial resolution (Federal Rule of Evidence 702 US 1973; Federal Rule of Evidence 703 US 1973). Those rules provide controls on the admission of expert-proved evidence such that certain "tests" must be met for the use of such expert-proved evidence.

FRE Rule 702 provides that:

> Federal Rule Evidence 702 (US) A witness who is qualified as an expert by knowledge, skill, experience, training, or education may testify in the form of an opinion or otherwise if:
> (a) the expert's scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue
> (b) the testimony is based on sufficient facts or data
> (c) the testimony is the product of reliable principles and methods; and
> (d) the expert has reliably applied the principles and methods to the facts of the case

FRE 703 provides that:

> An expert may base an opinion on facts or data in the case that the expert has been made aware of or personally observed. If experts in the particular field would reasonably rely on those kinds of facts or data in forming an opinion on the subject, they need not be admissible for the opinion to be admitted. But if the facts or data would otherwise be inadmissible, the proponent of the opinion may disclose them to the jury only if their probative value in helping the jury evaluate the opinion substantially outweighs their prejudicial effect.

These requirements may be difficult to meet, especially for unsupervised or semi-supervised artificial intelligence and machine learning systems where, in part, the system is effectively programming itself to do certain tasks.

For electronic systems that have provided evidence in legal fora in the past, such as photographic and video cameras, some fora permit inferential validation by testimony that the system is regularly used, was used as it was meant to be used, and operated properly. The burden is shifted back onto those challenging the evidence to show that the system was not working properly and cannot be relied upon reliable evidence.

This may be difficult where the systems cannot be evaluated due to their complexity and occult operations. Testimony that the system has been in use and operates properly

may not be sufficient especially where digital imagery can be easily manipulated; deep fake systems are designed for difficult-to-detect manipulation. Such systems may not be available for evaluation and study as to aid in the detection of their use.

The growth in number and variety of sensing devices connected via Internet technologies greatly expands data generated on everyone and available for analysis. Validating the reliability of this data collected becomes even more important. An inventory of the ways in which data can be sensed and transmitted to central storage and for analysis demonstrates the broad and ubiquitous means by which human activity can be surveilled. A partial inventory includes these sensing systems, whether individual or as part of larger data collection systems and shows the plethora of devices for capturing aspects of the lives of others:

- video surveillance and identification
- video surveillance and behavioural inference
- audio surveillance (ShotSpotter for neighbourhood gunfire)
- license plate readers
- automobile networks and toll scans
- Smart City technology
- Cell Site Location Information (CSLI)
- other sensor networks
- RFID
- home IOT devices
- out of home IOT devices
- other sensors that come with evolving technology

Each of these systems generates information of varying types, all accompanied by various types of metadata, such as time and place, which allow additional inferences regarding a data subject's activities. Video information can be fed into facial recognition systems that, along with the time and location data of the video, can place a person in a particular place at a particular time. Cell site location information (CSLI) can place an individual proximate to illegal or embarrassing activity, with varying degrees of accuracy and precision.

Law enforcement agencies have sought CSLI information that might connect individuals in a state that prohibits pregnancy termination to pregnancy termination clinics in another state. Audio surveillance for gunshots as to time and location can direct of additional police resources to a particular location; those police resources have been informed of possible "shots fired" and may respond accordingly, with guns drawn. Data analytics against image technology can "identify" suspects and possible offenders, but an error rate in the analytics may lead to erroneous mortal outcomes for predictive policing.

License plate readers, service tagging devices such as for toll roads, and "Smart Cities" implementations of sensing devices can support services ranging from traffic management to public health. They provide a skein of data for both directly managing related services. But such data profiles can be used, alone or in conjunction with cross matching against other data, to create profiles on individuals or classes of individuals. The growth of the Internet of Things and all of the sensing and data collection devices associated with such

systems, within homes, within businesses and in the public world, create another bounty and wealth of data that can be accessed and collected for analysis.

Where that information is proximate to potential illegal activities, it may create the foundation for further state surveillance and investigation of an individual, regardless of the guilt or innocence of that individual. It also creates the ability to profile and make inferences on the lives of individuals in unprecedented ways. This may or may not be within the purview of the state but, nonetheless, puts immense power in the hands of state actors to use. Such uses may not be in the interest of the individual but that of others seeking to manipulate and impact her.

The domains of different nodes for the creation and transmission of data present offer guidance for regulation and engineering to preserve privacy. Regulation may involve some or all of these information spaces as to protect privacy, as shown in Figure 1.

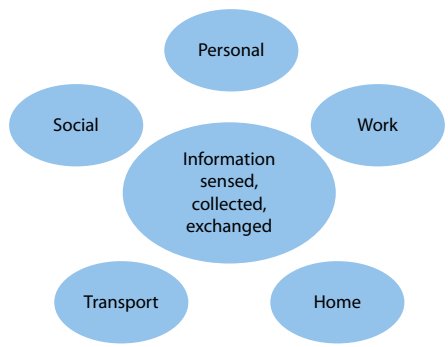

Figure 1.
Data nodes of activities and services

Source: Losavio et al., 2018

4. How data is collected and stored for surveillance

After the sensing and generation of data, it is available for use. While that use may be local at the sensing device, there is greater utility in the collection of large amounts of data for more sophisticated analysis and broader use. The Internet of Things paradigm extends from the local generation of data to its transmission and collection, often through Internet technologies. After transmission there is the collection of the data various points for use. The evolving topology of the Internet of Things now includes Edge computing, Fog computing and Cloud computing. In Edge computing data is collected close to the point of generation for use and analysis, either locally or for downstream transmission. Fog computing are those systems connected in the space between the Edge/sensors and the Cloud computing environment; Fog computing allows for regional aggregation and analysis of data and use. The Cloud metaphor for computing addresses endpoint services of massive storage and computational power connected globally to provide services and data processing results as needed.

Whether through Internet systems or through private networks, regulation may be applied within the various space of the topology. The GDPR limits data transmission across national boundaries, defining the facts that indicate relevant regulation of data transmission, analysis and usage within that apology can be controlled.

The topology, or structure by which data is collected, transmitted and stored encompasses a wide variety of systems. The systems may be mobile or stationary, domestic transnational, proprietary/third party, governmental or open source. All of these areas may be subject to regulation, including controller transmission and exchange of data.

Limits on government access to data, including third party data can have similar protective benefits through those limitations. Such protective regulation includes these as to reinstate the "practical obscurity" that allowed privacy over time, similar to the "right to be forgotten".[12] An expansion of the notion of "practical obscurity" would likely provide for redaction of the information in records as a means of protection for privacy. But the ability of electronic databases to tag particular data as to obscure it from certain search requests, especially those from a particular jurisdiction, may let such delisting be sufficient subject to damages where it fails to conceal that information.

Examples of such protective limits and practices may include:

- government limits on data exchange
- First Amendment (U.S.) freedom of speech limits on regulation of data exchange
- GDPR limits on data exchange
- voluntary limits on data exchange
- personal responsibility for data generation, data transmission data exchange

Similarly, protections are seen in Article 17 of the GDPR:

1. The data subject shall have the right to obtain from the controller the erasure of personal data concerning him or her without undue delay and the controller shall have the obligation to erase personal data without undue delay where one of the following grounds applies:
   a) the personal data are no longer necessary in relation to the purposes for which they were collected or otherwise processed;
   b) the data subject withdraws consent on which the processing is based according to point (a) of Article 6(1), or point (a) of Article 9(2), and where there is no other legal ground for the processing;
   c) the data subject objects to the processing pursuant to Article 21(1) and there are no overriding legitimate grounds for the processing, or the data subject objects to the processing pursuant to Article 21(2);
   d) the personal data have been unlawfully processed;
   e) the personal data have to be erased for compliance with a legal obligation in Union or Member State law to which the controller is subject;
   f) for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) in so far as

[12] *United States Department of Justice v. Reporters Committee for Freedom of Press*, 489 U.S. 749 (1989).

> the right referred to in paragraph 1 is likely to render impossible or seriously impair the achievement of the objectives of that processing; or
>
> g) for the establishment, exercise or defence of legal claims.

The means to meet compliance with these rules by search engines was detailed in *GC and Others* EU:C:2019:773.[13] The jurisdictional scope of such rules was set out in *Google v. CNIL* EU:C:2019:772.[14]

Personal responsibility for data generation, data transmission and data exchange offer another means of protection. An individual's care in maintaining privacy may preserve that privacy, but only if attention is paid to efforts to obtain a person's agreement to surrender that privacy. All manner of contracts contain specifications for the release of private data. A person – a data subject – may "opt out" of such disclosures, but often that requires a positive act, even if only checking a box, to invoke that protection.

5. How is that data accessed, analysed and used for surveillance?

The endpoint of data creation, transmission and collection is its use. The growth of powerful analytics, especially that denoted "Artificial Intelligence (AI)", make for new concerns about the power of government surveillance. The "practical obscurity" of individual data in large data collections is no longer possible given the power of analytics to sift through trillions of bytes of data quickly and, by analysis, both find otherwise obscure bits of information but also process that information to find patterns and relationships. That analysis can produce conclusions and inferences that would otherwise have been impossible in the past without the allocation of immense resources.

These powers of analytics represent new and dangerous possibilities for the use and misuse of the systems for government surveillance. Examples of the immense analytical power now available include generative AI and Computer Vision/Facial Recognition. These offer exceptional power for government surveillance to promote public safety. The GDPR data processing regulations offer some control over the ways in which these systems apply, although some worry it is not a sufficient control over such systems (Mitrou, 2018).

The European Parliament adopted the EU Artificial Intelligence Act to provide a broad regulatory framework for AI. That AI Act banned some AI applications, albeit with specifically delineated exceptions for law enforcement, and strict regulation of "high-risk systems". The co-rapporteur for the Internal Market Committee noted:

> We finally have the world's first binding law on artificial intelligence, to reduce risks, create opportunities, combat discrimination, and bring transparency. Thanks to Parliament, unacceptable AI practices will be banned in Europe and the rights of workers and citizens will be

[13] Case C-136/17 *GC and Others* EU:C:2019:773. See the text of the decision at https://doi.org/10.1093/grurint/ikaa003. *GC* outlines how search engines implement the right to be forgotten in their databases.

[14] Case C-507/17 *Google v. CNIL* EU:C:2019:772. See the text of the decision at https://doi.org/10.1093/grurint/ikaa004. CNIL sets out the territorial coverage of the right to be forgotten as to define the obligations of data holders to implement that.

> protected. The AI Office will now be set up to support companies to start complying with the rules before they enter into force. We ensured that human beings and European values are at the very centre of AI's development.

Such restrictions as set out in the EU AI Act will limit how data and AI are used, albeit, with greater leeway for law enforcement pursuing public safety via such systems.

For the United States there is yet little regulation to date beyond common law principles and statutes relating to personal injury in violations of civil rights. A U.S. Artificial Intelligence regulatory statute is pending in the U.S. Congress. Other nations, such as the People's Republic of China and the Federative Republic of Brazil, have implemented initial regulatory statutes while pursuing further legislation. The United Kingdom has set out its own framework from which to begin its process of AI regulation.

This rush to build regulatory structures parallels the rapid growth in benefits and dangers from such systems. The concerns regarding AI mediated facial recognition systems for government surveillance derived from concerns as to reliability of the analysis and its subsequent use. Errors, whether false positives or false negatives, undermine the reliability of these systems, particularly as to gender and ethnicity where individuals from related groups may not have been adequately represented in the data training set (Axon, 2019). False negatives lead to a possibly guilty person escaping; false positives lead to an innocent person becoming the focus of police powers.

Regulation in this exceptionally important area of AI-mediated government surveillance can be problematic given the early stages of these technologies, particularly as to undeveloped standards for testing and validating these systems. It is a problem of not knowing what we don't know. This reflects some of the underlying problems in AI generally. The risks of injury from these problems and defects is especially high in government surveillance systems.

To assure the best outcomes from the use of Artificial Intelligence systems, those systems must be:

- transparent, in that their operations may be open to inspection and review
- explainable, in that the designers and operators of these systems understand those operations and can explain how the system produces the results it produces and the reliability that can be placed on those results
- accountable, in that where there is injury that results from these systems the responsibility for them can be allocated

Yet these may be elusive for a variety of reasons.

AI transparency may be hindered by the sheer complexity of those systems, especially where unsupervised self-learning algorithms have devised the operations beyond human coding. Those AI systems may also be hindered by efforts to protect proprietary design within the system.

AI "explainability" as to how an AI system operates is made difficult by the different ways in which the system may operate, especially where that system is trained to teach itself against large databases of information. Unsupervised machine learning algorithms are trained against large databases of information and code themselves to develop

result-producing algorithms from their analysis of that data. The resulted algorithms may not be human-mediated and may require extensive back analysis to understand their operations. Without that there may be no way to explain exactly what the system does; all that can be seen are the results of system operation. Waiting to evaluate those results after operation may lead to injuries of others as that evaluation is done.

AI accountability for damage done by AI mediated systems is itself hindered by the problems with the lack of transparency in operations and that the systems and their operations cannot be explained due to the way in which these systems developed. This makes the attribution of causation, whether to the developers/programmers, vendors or users, much more problematic. Further limitations on accountability may be found within the end-user license agreements that may accompany the transfer of ownership or license of an AI system to the user or third party.

6. The human-in-the-loop factor for surveillance: Additional protection or additional abuse

A factor increasingly seen as vital in the use of AI, and of particular importance when used for government surveillance, is the notion of the "human-in-the-loop" factor, that there is a human-mediating element to the use of AI systems. AI results should be subject to human review of the outcomes as to provide a critical limitation on damage from erroneous outputs. But the reliability of this component may vary, depending on the very competence or integrity of the particular human receiving a particular AI output in the surveillance of a person, community or nation. The human-in-the-loop aspect of AI systems for government surveillance raises additional concerns beyond making the system more reliable. The very integrity of that human becomes an issue as to proper vetting of the AI results as to rely on them.

The ethics review board of the Axon law enforcement technology vendor was charged to evaluate facial recognition technology and its utility for public safety. It specifically noted a problem where such system could be modified by system users; the risks of such manipulation within law enforcement were deemed further reason facial recognition technologies should not yet be deployed. But human engagement can have an impact on implementation of AI outputs; the State of Michigan's "robo-judge" system for determining unemployment insurance fraud went from a 90+% false positive rate to a 50+/–% false positive rate once human reviewers were inserted into the process.[15] Encouraging, but not sufficient. In the evaluation of LASER and PREDPOL analytical policing systems, the evaluators found errors relating to the lack of consistency in the human component as to data collection, analysis and implementation.[16]

[15] *Cahoo, et al. v. SAS Analytics Inc., et al.* ___ F.3d ___ (6th Cir. 2019) (US).

[16] Report of the Inspector General-LAPD, *Review of Selected Los Angeles Police Department Data-Driven Policing Strategies,* March, 2019.

One indication of the importance of the human-in-the-loop is the notion of "prompt engineering" in support of the use of Large Language Models (LLM) (Saravia, 2022). These LLMs are essential to the training of GPT systems.

This is the idea that most effective use of generative AI and Large Language Models (LLM) such as ChatGPT is where the human can devise the most effective and efficient inputs to the system as well as be able to judge the reliability of the outputs. This will require a deeper understanding by humans of system operations and linguistic concepts as to build effective AI prompts. This is a nascent area of skill and ability that must be developed to provide better assurance of reliable outcomes.

Human engagement and attention to these operations can support reliability and accountability. It may be the final bulwark protecting subjects of AI analysis. But that assumes the competency of the humans involved. And their integrity and honesty.

The final implementation for reliability and safety in AI-mediated government surveillance will be assuring competency and integrity of the human component. That may require:

- testing and certification of human operator as to competency and integrity
- training of human operators to ensure competency and integrity of use
- a compliance regime to guide and assure proper use of AI systems that includes logging of operations and auditing of compliance by all elements in the system

Without this it will be much more difficult to know until *after* injury happens that the system was used improperly.

7. Conclusion – Balancing order v. liberty

These issues relating to AI reliability go beyond government surveillance and predictive policing into all implementations of these systems and the risks they pose (Gianni et al., 2022). This has led to nations seeking strategic approaches to the implementation of AI in public administration as to mitigate those risks (Dutton, 2018; Berryhill et al., 2019; Misuraca & Van Noordt, 2020). The EU Artificial Intelligence Act is an important first step in building a coherent regulatory framework that both promotes the amazing abilities of this technology and the protection of the rights of people who may be threatened by it.

Assuring proper use of AI-mediated government surveillance includes technical, legal and end-user issues that are part of AI governance. From design to implementation to compliance review, these are systems within which well-drafted mandates best assure the reliability and propriety of the systems and their use. Failure to create an effective system of governance that holds everyone to account in the chain of operations – from data to AI to human – destroys accountability and any incentive for fair and just systems.

How the state engages in this process of regulation is critical. There must be government control and government restraint and ethical use as to implementation, compliance and governance (Winfield & Jirotka, 2018). Such engagement may need to be more responsive and timely given the rapid change in the technology (Wallach & Marchant,

2018). The broad use of the Internet of Things, such as RFID chips to tag so many things and the ability of CSLI to turn any cellphone in a tracking device, and the potential for their misuse, demonstrate the risks of ever-novel technologies.

This governance must be informed, intelligent and flexible. The uproar over AI systems, particularly generative pre-trained transformer systems like ChatGPT, makes this an immediate concern. There have been calls to freeze research and development on AI systems, a solution that will not satisfy nor protect anyone. Rather, this gives an advantage to research in regimes that have less scruples about these matters and may give them a competitive advantage in crucial areas. These include areas of public safety and national security, as the use of AI systems in cyberattacks is an increasingly sophisticated and cost-effective means to break down, infiltrate and exploit the cyber systems vital to contemporary life.

The engagement of public safety, law and AI specialists can promote this. Review and policy boards on AI are increasingly used to provide flexible and informed governance while promoting innovation. The failure to work together may lead to the terrible outcomes we fear. We must not let that happen.

References

Axon Enterprise, Inc. (2019). *First Report of the Axon AI & Policing Technology Ethics Board.* Online: https://shorturl.at/kyEHR

Bensinger, K. & Chang, A. (2013, April 20). Boston Bombings: Social Media Spirals out of Control. Los Angeles Times. Online: https://shorturl.at/nstEU

Berryhill, J., Heang, K. K., Clogher, R. & McBride, K. (2019). Hello, World: Artificial Intelligence and Its Use in the Public Sector. *OECD Working Papers on Public Governance,* (36). Online: https://doi.org/10.1787/726fd39d-en

Campbell, D. (2001, October 17). US Buys Up All Satellite War Images. *Manchester Guardian*.

CBS News (2013, April 24). Social Media and the Search for the Boston Bombing Suspects. *CBS News*. Online: https://shorturl.at/gJQTU

Connors, B. & Zauderer, A. (2013, January 2). Police Use Facebook to ID Waterford Robbery Suspects. *NBC Connecticut*. Online: https://shorturl.at/hmW34

Davidson, R. (2019, August 8). Automated Threat Detection and the Future of Policing. *FBI Law Enforcement Bulletin*. Online: https://shorturl.at/erPZ2

Deeks, A. S. (2018). Predicting Enemies. *Virginia Law Review, 104*(8), 1529–1592. Online: https://shorturl.at/lxyDM

Dutton, T. (2018, December 6). Building an AI World: Report on National and Regional Strategies. *CIFAR*. Online: https://shorturl.at/dlFJ3

Dyer, J. (2013, April 27). Social Media and the Boston Bombings. *BBC Radio.* Online: https://shorturl.at/gmoC0

Gianni, R., Lehtinen, S. & Nieminen, M. (2022). Governance of Responsible AI: From Ethical Guidelines to Cooperative Policies. *Frontiers in Computer Science,* 4. Online: https://doi.org/10.3389/fcomp.2022.873437

International Association of Chiefs of Police (2011). *Survey of Law Enforcement's Use of Social Media Tools*.

Losavio, J. D. & Losavio, M. M. (2014). Prosecution and Social Media. In C. D. Marcum & G. E. Higgins (Eds.), *Social Networking as a Criminal Enterprise* (pp. 197–220). Taylor & Francis. Online: https://shorturl.at/tBF69

Losavio, M., Elmaghraby, A. & Losavio, A. (2018). Ubiquitous Networks, Ubiquitous Sensors: Issues of Security, Reliability and Privacy in the Internet of Things. In N. Boudriga, M. S. Alouini, S. Rekhis, E. Sabir & S. Pollin (Eds.), *Ubiquitous Networking. UNet 2018.* Lecture Notes in Computer Science (LNCS), 11277. Springer. Online: https://doi.org/10.1007/978-3-030-02849-7_30

Misuraca, G. & Van Noordt, C. (2020). *AI Watch. Artificial Intelligence in Public Services.* Publications Office of the European Union. Online: https://doi.org/10.2760/039619

Mitrou, L. (2018). *Data Protection, Artificial Intelligence and Cognitive Services: Is the General Data Protection Regulation (GDPR) 'Artificial Intelligence-Proof'?* Online: https://dx.doi.org/10.2139/ssrn.3386914

Perry, W. L., McInnis, B., Price, C. C., Smith, S. & Hollywood, J. S. (2013). *Predictive Policing. The Role of Crime Forecasting in Law Enforcement Operations.* RAND Corporation. Online: https://doi.org/10.7249/RR233

Saravia, E. (2022). Prompt Engineering Guide. *GitHub*. Online: https://github.com/dair-ai/Prompt-Engineering-Guide

Stuart, R. D. (2013, February 5). Social Media: Establishing Criteria for Law Enforcement Use. *FBI Law Enforcement Bulletin*. Online: https://shorturl.at/intzO

Wallach, W. & Marchant, G. E. (2018). *An Agile Ethical/Legal Model for the International and National Governance of AI and Robotics.* Association for the Advancement of Artificial Intelligence. Online: https://shorturl.at/deHT7

Winfield, A. F. & Jirotka, M. (2018). Ethical Governance Is Essential in Building Trust in Robotics and Artificial Intelligence. *Philosophical Transactions of the Royal Society A Mathematical, Physical and Engineering Sciences, 376*(2133). Online: https://doi.org/10.1098/rsta.2018.0085