12-2020

Healthcare Regulation and Governance: Big Data Analytics and Healthcare Data Protection

Xuejuan Zhang

Healthcare Regulation and Governance:

Big Data Analytics and Healthcare Data Protection

Xuejuan Zhang

Several definitions of "big data" have been suggested in the literature as efforts have been made by many stakeholders to understand this new field. For this paper, the consensus definition proposed by Grady (2019) in a report published by the National Institute of Standards and Technology, a division of the U. S. Department of Commerce will be used. Grady (2019) wrote, "Big Data is a term used to describe the large amount of data in the networked, digitized, sensor-laden, information-driven world (p. iii)."

The characteristics of big data that force new architectures can be summarized by "4 Vs" which refer to Volume (i.e., the size of the dataset); Velocity (i.e., rate of flow); Variety (i.e., data from multiple repositories); and Variability (i.e., the change in velocity or structure).[1]

Grady (2019) argues that the 4 Vs are fundamental drivers dictating the overall design of a Big Data system resulting in different data system architectures or different analytics life cycle process orderings to achieve desired performance and cost-efficiency. Cost-effective data collection, storage, and processing have enabled users across various industries to manage the size, speed, and complexity of Big Data. In a more digitally connected society and economy, technological advancements differentiate in advanced analytical tools.

Big Data in Healthcare

Healthcare in the U.S. is a complex ecosystem consisting of various stakeholders including the following groups (1) public and private healthcare industry members consisting of patients, clinicians, private payers (i.e., insurance companies), and researchers; (2) broad healthcare sector members that expand to the business associates of the above-defined private healthcare industry and the public health providers and financiers; and (3) an even more broadly defined healthcare ecosystem that includes a broad base of general consumers who have needs for improving personal health and well-being and utilize the services provided by either public or private healthcare industry, and (4) those public or private industries or entities who contribute to the delivery of such services. Figure 1 presents a stakeholder view of the Healthcare Ecosystem, and Figure 2 presents a relationship diagram among the stakeholders.

[1] Big data do not have qualitative characteristics. There is no consideration, for example, of confusion about meaning, trust about accuracy (reliability or validity), or cultural relevance. This suggests that none of these is considered when the design for performance or efficiency is created.

Figure 1. System of Health Care Stakeholders

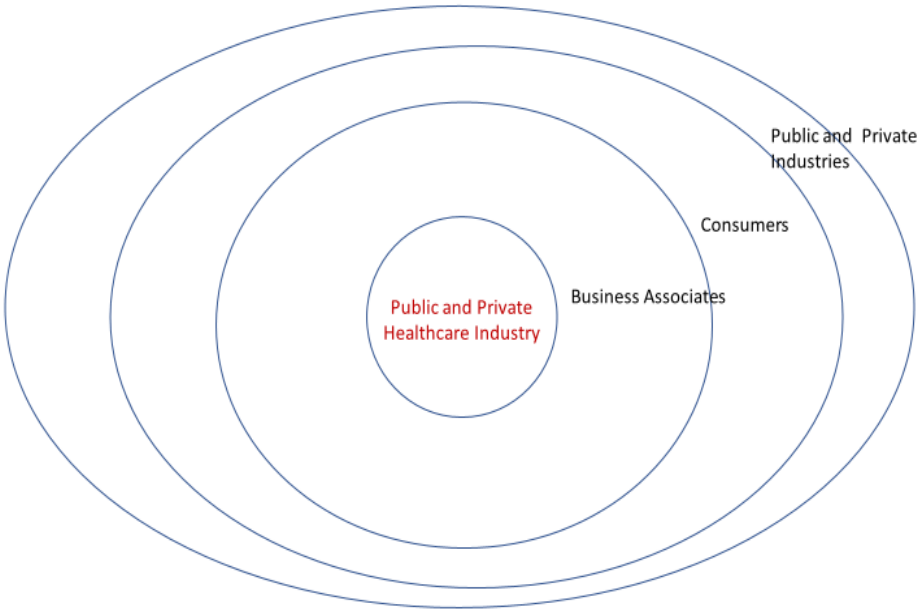
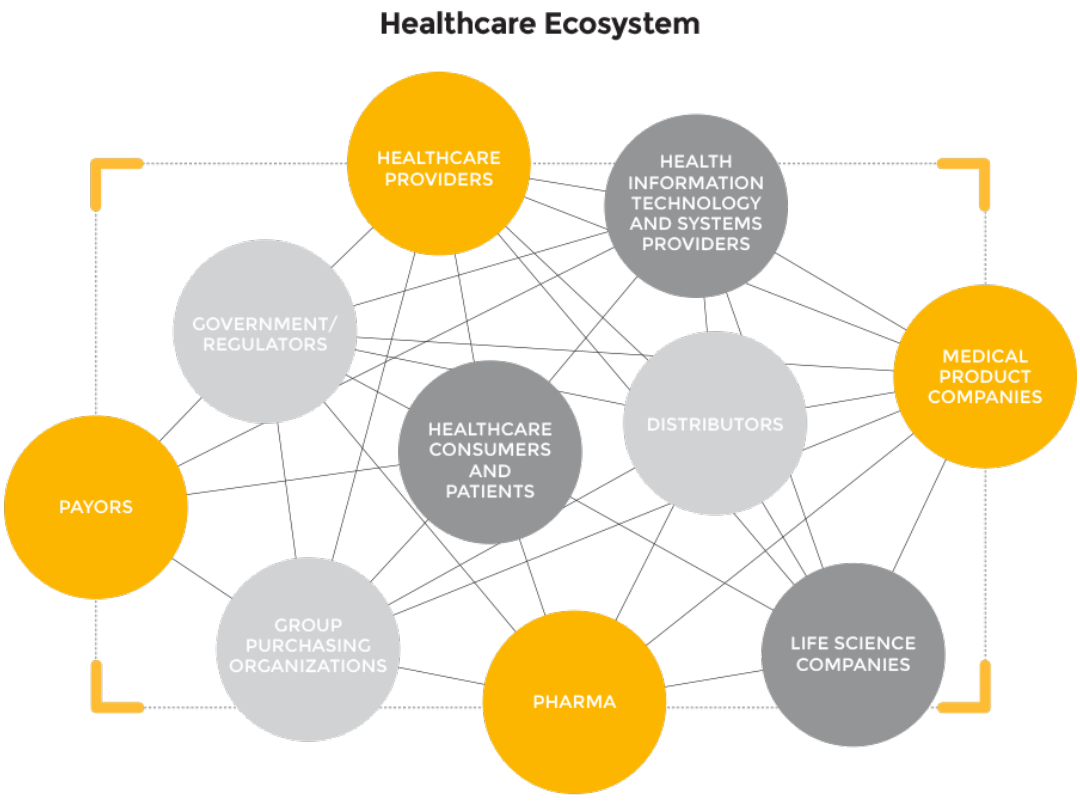


Figure 2. Stakeholder Relationship Diagram (Source: https://www.moveo.com/what-we-do/healthcare-ecosystem/)



The healthcare ecosystem recognizes that broadly defined healthcare includes not only a person's episodic medical encounters such as medical and/or care history but also comprehensive health and lifestyle information. For instance, the manufacturers of personal wearables (e.g., Apple Watch, Fitbit wrist band, etc.) and their partners such as Nike have developed hardware (Apple Watch) and software (e.g., the application of Nike Run Club that can be installed on Apple Watch and Apple Smartphone) to help people to track their health conditions and health-related activities such as diet, mental health, and recreation.

These interdependencies among elements have led to emergent outcomes that become an unignorable force, especially from the data perspective, because the advancement of technology has enabled real-time personal health data and behavioral data production and collection in vast volume and speed. And this massive health and behavioral data generation fall outside the scope of the traditional narrowly defined healthcare industry.

In the past three decades, the significant digitalization of data in the U.S. healthcare sector is represented by the development of Electronic Health Records (EHRs), the systemized collection of patient and population electronically stored information. The paper-based health record, referred to as an Electronic Medical Record (EMR) is a term often used interchangeably with EHR. But EHR is a more longitudinal collection of health information of patients or populations, while an EMR is created by certain providers for specific medical encounters, which can serve as the data source for an EHR.

The earliest EHR was developed in 1971 and gained popularity by 1992 (Evans, 2016). EHRs were initially used for claim processing and a means for document capture and later for clinical decision support (CDS). There are at least two significant changes in the evolution of EHRs in the U.S. First, personal data are integrated into an EHR, such as mental and behavioral data, family history, and non-medical data of key life events including data that are external to the healthcare provider (test results from labs). Second, the EHR users are fast expanding from clinic to primary care physicians, hospitals, insurance companies, patients, and nursing homes. This means that EHRs are increasingly used with online medical information and decision-making tools, which have changed the dynamics of the patient-clinic interaction through clinician-patient email, virtual consults, and telemedicine (Evans, 2016).

Through utilizing different sources of data (i.e., medical, lifestyle, financial data), big data and the application of methods to examine and interpret referred to as big data analytics may not only help individuals to improve their health and prevent them from seeking care from healthcare providers but also help to build a dynamic learning process that can identify effective treatments, drugs, and public health interventions that improve the efficiency of the overall healthcare system. However, to take full advantage of big data requires that data from different sources can be collected, transferred, and integrated with advanced analytical tools such as artificial intelligence and machine learning effectively. However, this is a challenge because there exists a

conflict between the public's concern about data privacy and the required data liquidity to reach the full potential of big data in healthcare.

Regulatory Framework of Healthcare Data in the U.S.

The development of big data, which enables the collection and transmission of personal data in vast quantities, has raised the public's concerns and drew the attention of regulators. For instance, the Organization for Economic Co-operation and Development (OECD), an intergovernmental economic organization that provides advice on public policies and international standard-setting, updated its guidelines regarding protecting the privacy of personal data. OCED (2013) lays out the basic principles of both national and international applications, including a set of principles (Table 1) of collection limitation, purpose specification, use limitation, individual participation, and accountability, collectively known as the Fair Information Practice Principles (FIPPs).

Table 1: OECD Fair Information Practice Principles

Basic Principles	**Explanation**
Collection limitation	There should be limits to collecting personal data and any such data should be obtained by lawful and fair means with the knowledge or consent of the data subject.
Purpose specification	The purposes for which personal data are collected should be specified not later than at the time of data collection and the subsequent use limited to the fulfillment of those purposes.
Use limitation	Personal data should not be disclosed, made available, or otherwise used for purposes other than those specified except 1) with the consent of the data subject; or 2) by the authority of law.
Security Safeguards	Personal data should be protected by reasonable security safeguards against such risks as loss or unauthorized access, destruction, use, modification, or data disclosure.
Openness	There should be a general policy of openness about developments, practices, and procedures with respect to personal data.
Individual participation	An individual should have the right to confirm whether or not the data controller has data relating to him; to have data relating to him within reasonable time cost; to be given reasons if a request of obtaining data relating to him to be denied.
Accountability	A data controller should be accountable for complying with measures which give effect to the principles stated.

A White House report (Executive Office of the President, 2014) on big data highlighted the mounting pressures on traditional privacy protection measures (i.e., FIPPs). This report acknowledges the complex legal landscape regarding patient information and privacy and highlights that the current federal law (i.e., HIPPA 1996) and other privacy laws regulating the collection of health-related information may not meet consumers' expectations of privacy. The President's Council of Advisors on Science and Technology (PCAST, 2014) reinforces the pressure big data places on the FIPPs from a technical perspective. The Health IT Policy Committee (HITPC) Privacy and Security Workgroup released a report (HITPC, 2015) that

provides discussions and recommendations about privacy and security concerns, potentially harmful uses of big data in healthcare, and potential gaps in privacy and security protection. As of this writing, Congress is still debating about a national data privacy law to reduce the risk of consumer data (Healthitsecurity, 2019).

Despite the increased interest in data protection, the legal framework governing the privacy of personal data is complex, lacks uniformity at the federal level, and may be best described as "patchwork" (Mulligan, Linebaugh, & W. C. Freeman, 2019). That is, there are several data protection statutes at the federal statutory level which regulate certain industries and subcategories of data. These laws vary considerably in their purpose and scope, i.e., governed entities and data protection requirements. Table 2 provides a summary of major federal data protection laws. For instance, the Gramm-Leach-Bliley Act (GLBA) 67 imposes several data protection obligations on financial institutions.

Table 2: Major Federal Data Protection Laws

Federal Law	**Information Protected**	**Covered Person**	**Nature of Regulation**
Gramm-Leach-Bliley Act (GLBA)	Nonpublic personal information (NPI)	Financial institutions	Consumer opt-out requirement for data sharing; Consumer disclosure and data security requirements
Fair Credit Reporting Act (FCRA)	Consumer reports	Credit Reporting Agencies (CRAs), furnishers of information to CRAs, and users of consumer reports issued by CRAs	Accuracy and use requirements for consumer reports; Consumer disclosure requirements
The Communications Act	Customer proprietary network information (CPNI) Personally identifiable information (PII)	Common carriers Cable operators and satellite carriers	Consumer consent requirement for data sharing; Consumer disclosure and data security requirements
Federal Trade Commission Act (FTC Act)	n/a	All persons or commercial entities other than common carriers, certain financial institutions, and nonprofits	Data privacy and security policies and practices must not be "unfair or deceptive"
Health Insurance Portability and Accountability Act (HIPAA)	Protected health information (PHI)	Healthcare providers, health plans, and health care clearinghouses	Consumer consent requirement for data sharing; Consumer disclosure, Data security and data breach disclosure requirements

Note: this table is adapted from the Appendix of Mulligan, Linebaugh, & W. C. Freeman (2019).

Currently, the legal framework that protects health information privacy in the U.S. is governed by federal and state laws. At the federal level, the principal law is the Health

Information Portability and Accountability Act of 1996 (HIPPA). HIPPA governs "protected health information" (PHI), which is individually identifiable information about an individual's care, health condition, or payment for care. Under HIPPA, the Privacy Rule applies to "regulated entities," including "covered entities" (i.e., health plans, health-care clearinghouses, and most healthcare providers), and their "business associates" (i.e., entities having access to or using PHI when performing specified functions or services for the covered entity).

Each state defines its own privacy framework, which usually governs the same entities, activities, and information as the federal laws. State laws often provide enhanced protections for sensitive information or vulnerable population (Thorpe & Gray, 2015). In particular, a State law is "contrary" to the HIPAA Privacy Rule if it would be impossible for a covered entity to comply with both the State law and the Federal Privacy Rule requirements, or if the State law is an obstacle to accomplishing the full purposes and objectives of the Administrative Simplification provisions of HIPAA.

With certain exceptions, the Privacy Rule preempts "contrary" State laws.[2] More broadly, according to the National Conference of State Legislatures (NCSL), more than half the States considered or introduced consumer data privacy legislation (Greenberg, 2019), and 31 states enacted cybersecurity-related legislation in 2019 (NCSL, 2019). For example, the California Consumer Privacy Act of 2018 (CCPA) became effective on January 1, 2020. CCPA represents one of the broadest online privacy laws in the U.S. Additional legislation was introduced in 2020 to address the collection and use of biometric or facial recognition data by commercial entities. For instance, pending the California voters' approval, California Proposition 24 will further expand the state's consumer data privacy laws. It will limit businesses' use of "sensitive personal information," such as precise geolocation, health, and biometric information. It will establish a California Privacy Protection Agency to enforce and implement consumer privacy laws and impose administrative fines.[3] Many other states, such as Illinois, Connecticut, and New York are establishing new consumer data privacy rules to regulate the disclosure of personal information such as health and genetic testing information and to regulate data brokers collecting of personal data.

The U.S. regulatory approach of healthcare information can be characterized as a sectoral and downstream regulation approach (Terry, 2017). The sectoral approach means the entities and/or information governed by HIPPA are very narrowly defined. Different industries or sectors may have their own privacy regulations that govern the information for the same patient but not health-related (such as genomic information and financial information). The lifecycle or value-chain of data may be characterized as a linear sequence of collection, processing, storage and transfer, and final uses.

---

[2] Source: https://www.hhs.gov/sites/default/files/adminsimpregtext.pdf?language=es

[3] Source: https://www.ncsl.org/research/telecommunications-and-information-technology/2020-consumer-data-privacy-legislation637290470.aspx

Two concepts related to information protection are "privacy" and "confidentiality." "Privacy" provides protections against unauthorized data collection, and "confidentiality" provides protections against data controllers' disclosures of collected data. Terry (2017) argues that HIPPA is more a downstream approach that may inadequately address the data collection. For instance, data collections outside the traditionally defined healthcare industry are largely ignored. The recent debate highlights that as the value-chain of big data gets more complicated, the regulation of use (downstream) seems to be inadequate (Healthitsecurity, 2019).

The systems view of health care refers to very broadly defined stakeholders (in Figure 1) and government regulators (Figure 2) including entities outside the U.S. One example is the EU's General Data Protection Regulation (GDPR) that became effective on 25 May 2018. It regulates the processing by an individual, a company, or an organization of personal data relating to individuals in the EU. In particular, health data are considered sensitive data under the GDPR Article 9 and the processing thereof can, therefore, only take place under strict requirements. GDPR is relevant to US regulations because the set of data protection rules described applies to all companies operating in the EU, wherever they are based including those in the U.S.

Regulations of Artificial Intelligence and Machine Learnings

Another of the interdependent elements of the health care system ecosystem (Figure 2) is health information technology, and specifically the use of advanced analytical tools such as artificial intelligence (AI) and machine learning (ML). The European Commission defines AI as "a collection of technologies that combined data, algorithm, and computing power" (European Commission, 2020). AI can also be integrated with hardware. In case of ML techniques, which constitute a subset of AI, algorithms are trained to infer certain patterns based on a set of data to determine the actions needed to achieve a given goal. Algorithms may continue to learn when in use (European Commission, 2020). AI has been promoted by advocates as having the potential to bring benefits to individuals, businesses, and society as a whole. For instance, a new generation of AI-backed products and services in healthcare (e.g., wearables and health apps) has had a significant impact on society.

At the same time, the broad adoption of AI also creates many legal and ethical challenges. Specific to the application of AI in healthcare, ethical challenges may include informed consent to use, safety and transparency, algorithmic fairness and biases, and data privacy (Gerke, Minssen & Cohen, 2020). Table 3 provide several examples of ethical challenges.

Table 3: Ethical Challenges Related to AI application in Healthcare

Ethical Issues	**Examples**
Informed consent to use	• A need to examine under what circumstances the principles of informed consent should be deployed in clinical AI space. E.g., the clinicians' responsibilities to educate the patients bout the AI and ML used by the system, the data inputs, and the possibility for bias. •Consumers' consent to AI health apps and chatbots that constantly update themselves along with user agreements.
Safety and transparency	•IBM Watson for Oncology software gave "unsafe and incorrect" recommendations for cancer treatment because the software was trained by "synthetic" cancer cases instead of real patient data. •AI developers shall be transparent about the data used and any shortcomings of the software because transparency can foster trust among stakeholders.
Algorithmic fairness and biases	•AI bears a risk for biases and discrimination due to the ML procedure and training data used. For instance, an AI-based clinical decision support software for skin cancer could provide inaccurate recommendation if it was predominantly trained by Caucasian patients. •AI developed for experts in resource-rich settings (e.g., high-income countries) will not necessarily recommend treatments that are accurate, safe, and fair in low-resource settings (e.g., low-income countries).
Data privacy	•Data used outside the doctor-patient relationship can negatively affect patients, such as impacting insurance premiums and job opportunities.

Because AI is one of the most critical applications of the data economy and can have a major impact on our society, governments and regulators have paid attention to the regulatory framework regarding AI. International organizations such as OECD, G20 and governments in Europe and Asia have proposed principles and regulatory frameworks in the past few years.

OECD adopted its Principles on Artificial Intelligence in May 2019. This is the first international standard agreed by governments. In June 2019, the G20 adopted human-centered AI Principles that draw from the OECD AI Principles in Ministerial Statement on Trade and Digital Economy.[4]The set of Principles promotes the trustworthiness of AI-based on its compliance with the law, human rights democratic values and diversity, transparency and responsible disclosure of adoption, human-centric approach for AI-based decision making, and responsible usage of AI by individuals and organizations. Those principles include the issues presented in Table 4.[5]

[4] Source: https://www.mofa.go.jp/files/000486596.pdf
[5] Retrieved from: https://www.oecd.org/going-digital/ai/principles/

Table 4: Principles of Trustworthiness

1) AI should benefit people and the planet by driving inclusive growth, sustainable development, and well-being. 2) AI systems should be designed in a way that respects the rule of law, human rights, democratic values, and diversity, and they should include appropriate safeguards – for example, enabling human intervention where necessary – to ensure a fair and just society. 3) There should be transparency and responsible disclosure around AI systems to ensure that people understand AI-based outcomes and challenge them. 4) AI systems must function in a robust, secure, and safe way throughout their life cycles and potential risks should be continually assessed and managed. 5) Organizations and individuals developing, deploying, or operating AI systems should be held accountable for their proper functioning in line with the above principles.

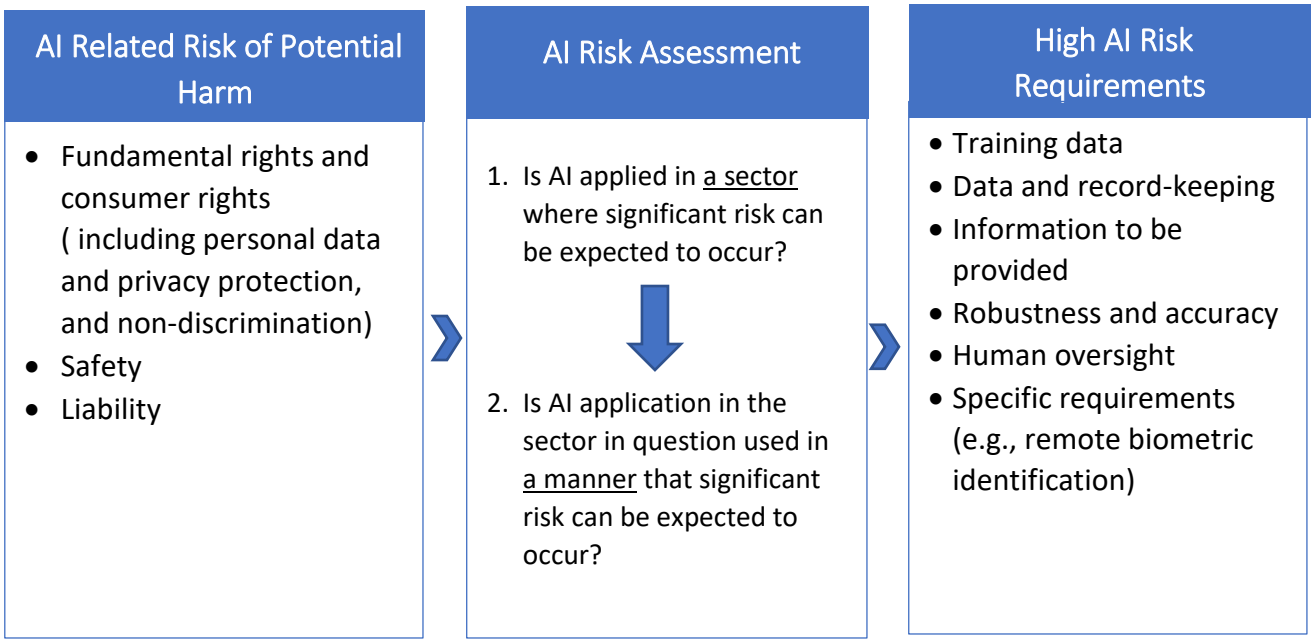
The European Union (EU) set its policy options for AI in February 2020. Similar to OECD principles, European AI is grounded in its values and fundamental rights, such as human dignity and privacy protection. EU's AI regulatory framework focuses on trustworthiness by building an "ecosystem of trust" along the whole value chain starting from research to the applications of AI in various scenarios. EU's AI regulatory framework recognizes that AI can cause harm. This harm might be both material (safety and health of individuals, including loss of life, damage to property) and immaterial (loss of privacy, limitations to the right of freedom of expression, human dignity, discrimination for instance in access to employment), and can relate to a wide variety of risks. Therefore, EU's AI regulatory framework adopts a risk-based approach to support that the regulatory intervention is proportionate (European Commission, 2020). Under this framework, an AI application should be considered high-risk, where it meets the following <u>two cumulative</u> criteria:

> (1) the AI application is employed in a sector where, given the characteristics of the activities typically undertaken, significant risks can be expected to occur.
>
> (2) the AI application in the sector in question is, in addition, used in such a manner that significant risks are likely to arise.

The first criterion addresses that the regulatory intervention is targeted to the areas where risks are deemed most likely to occur, such as healthcare, transport, energy, and parts of the public sector. The second criterion reflects the acknowledgment that not every use of AI in the selected sectors necessarily involves significant risks. For example, while healthcare generally may be a relevant sector, a flaw in the appointment scheduling system in a hospital will normally not pose risks of such significance as to justify legislative intervention. The assessment of the level of risk of a given use could be based on the impact on the affected parties (European Commission, 2020).

Several requirements would apply to high-risk AI applications only, including training data, data and record-keeping; information to be provided; robustness and accuracy; human oversight; with specific requirements for certain particular AI applications, such as those used for purposes such as remote biometric identification (European Commission, 2020). For instance, regarding human oversight, the regulatory framework (Figure 3) requires that the design of AI system must impose operational constraints of AI, the monitoring of AI consider the ability of human intervention in real-time, and the output of AI system that becomes effective immediately must be ensured with human intervention (e.g., human review) afterward.

Figure 3: EU Risk-based AI Model Framework



On January 23, 2019, Singapore implemented the first edition of the Model AI Governance framework (Model Framework), which was revised on January 21, 2020. In the Model Framework, there are two principles for responsible AI: (1) decisions made by AI should be explainable, transparent and fair; (2) AI solutions should be human-centric.

The Model Framework also highlights four areas for consideration: (1) internal governance structure and measures (including clear roles and responsibilities in organizations; SOPs to monitor and manage risks; staff training); (2) determining the level of human involvement in AI-augmented decision-making (including appropriate human involvement; minimizing the risk of harm to individuals); (3) operations management (including minimizing bias in data and model; risk-based approach to measures such as explainability, robustness and regular tuning); (4)

stakeholder interaction and communication (including making AI policies known to users; allowing users to provide feedback; making communications easy to understand).[6]

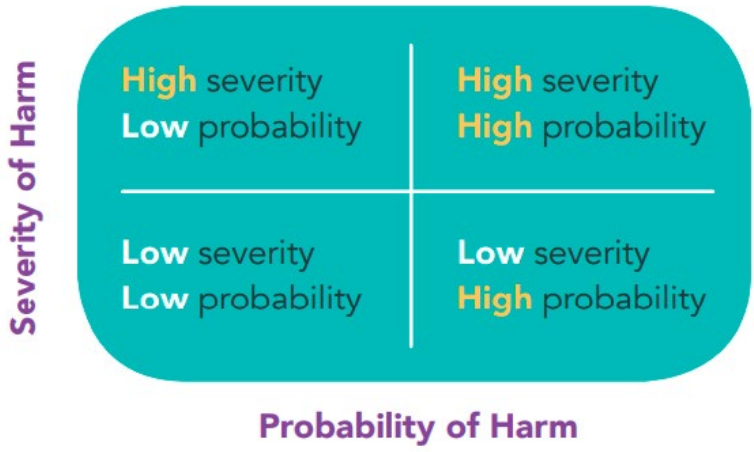
In general, depending on one's role in a decision-making process, the decision models can be classified into three types presented in Table 5.

Table 5: Decision Model and Human Intervention

Decision-making Type	Human's Role	AI's Role
Human-in-the-loop	Human oversight is active and involved in and interpretation and retains full control. Decisions cannot be exercised without affirmative actions by humans.	Only provides input or recommendations
Human-out-the-loop	No human oversight over the execution of decisions	AI has full control over the data analysis and decision-making without the option of human override
Human-over-the-loop or Human-on-the-loop	Human is in a monitoring or supervisory role, e.g., can adjust parameters during the execution of the algorithm	AI has certain control over the data analysis and decision-making

The Model Framework proposes a design framework (see Figure 4) to help organizations determine the level of human involvement required in AI-augmented decision-making. This design framework is structured along two axes: (a) probability; and (b) severity of harm to an individual (or organization) as a result of the decision made by an organization about that individual (or organization).

Figure 4: Risk Assessment Matrix for Harm due to AI



[6] Source: https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/SGModelAIGovFramework2.pdf

The definition of "harm" and the computation of probability and severity depend on the context and vary from sector to sector. For example, the harm associated with a wrong diagnosis of a patient's medical condition differs from a clothing store's wrong product recommendation for apparel. More importantly, "harm" is also a dynamic and temporally non-linear concept because the harmful results related to the utilization of AI in certain practices may not be perceived or understood immediately.

Complexity of Big Data in Healthcare

The outbreak of coronavirus (COVID-19) in 2020 and the subsequent efforts to contain this pandemic using technologies provide an example of the complexity of data privacy in the public health sector. According to WHO, the confirmed cases of COVID-19 infection and death worldwide on October 30, 2020, were 45.92 million and 1.19 million, respectively. To improve understanding of COVID-19 infectious diseases via social distancing and contact tracing, there has been a global wave of experiments using smartphones involving Google and Apple in April 2020. In Europe, 17 EU member states were using some type of mobile contact tracing apps.[7] Mobile apps are also used in Asia (e.g., Korea and China) and Middle East (e.g., Kuwait and Bahrain).

The United States federal response has taken a decentralized approach, with most COVID-19 contact-tracing apps being developed and rolled out by state governments (Figliola, 2020). Individual states such as North Dakota, South Dakota, Nevada, and Utah have independently deployed DCT apps by engaging the private sector. Most of the contact-tracing apps in the US are built on the Apple-Google protocol announced in April 2020. [8] The users must voluntarily download and opt into an appropriate state or regional tracing app as well as opt into the tracking feature in the smartphone operating system. Once enabled, a person's smartphone will exchange anonymous identifier beacon keys with nearby smartphones using the Bluetooth signal. The randomly generated identifier beacon key remains on a person's smartphone unless they report a positive COVID-19 diagnosis through the app. If one of the users later enters information (says) to the app that they have tested positive for COVID-19, the phone will upload the last 14 days of proximate contact data to a server, and those logged contacts will get alerts on their smartphones. For instance, the covered entity will include any entity or persona engaged in contact-tracing and exposure notification or develop tools for contact-tracing and exposure notification. Moreover, each bill will protect the specific data collected (such as geolocation data or any information linked or reasonably linked to any individual), besides personal health information.

[7] Source: https://ec.europa.eu/info/live-work-travel-eu/health/coronavirus-response/travel-during-coronavirus-pandemic/mobile-contact-tracing-apps-eu-member-states_en
[8] Source: https://covid19.apple.com/contacttracing

However, the experimentation of tracing apps does not meet its expectation. An article from *Washington Post* attributes the setbacks of tracing apps to the limited functions and people's distrust (Washington Post, 2020). For instance, the tracing apps only send alerts without information regarding when, where, and by whom people might have been exposed. Moreover, sensitive personal data, such as location data, that are key for tracing are not included in many apps in Western countries. In contrast, countries in Asia are willing to take more aggressive actions. For instance, the government in Hong Kong will not disclose confirmed COVID-19 cases that can be used to identify a specific person. But it will share the information regarding where they live and where they have been to in the past fourteen days with the public to take precautions. In the EU, health data are considered sensitive data under the GDPR (Article 9). However, the GDPR provides that one of the legal grounds for processing personal data is public interest in the area of public health. Aggregated statistical data that do not enable identifying the concerned natural persons (for instance aggregated location data) are not considered personal data, and therefore the GDPR does not apply.[9] Even though the COVID-19 tracing apps in Europe face similar problems that they have in the U.S., EU had set up an EU-wide system in October 2020 to ensure interoperability of COVID-19 contact tracing apps among member states.[10]

In the U.S., the pandemic of COVID-19 also put the flexibility and adaptability of HIPPA to a test in this public health emergency. In February 2020, the Office for Civil Rights (OCR) at the U.S. Department of Health and Human Services (HHS) provided a guideline regarding how patient health information (PHI) could be permissibly shared in compliance with HIPAA in the event of an outbreak of infectious disease or other emergencies. For instance, Health care providers may share PHI with anyone to prevent or lessen a serious and imminent threat to public health and safety (Shah & Hedgeman, 2020). However, in practice, it appears that in the U.S., neither the private sector nor the public sector takes advantage of those guidelines in tracing and containing COVID-19 possibly due to privacy concerns.

In response to the pandemic and the need for contact-tracing, as of June 26, 2020, data privacy bills had been introduced to address the privacy issues related to digital contact-tracing and exposure notification (Gaffney, 2020), including:

1. the COVID-19 Consumer Data Protection Act of 2020 (CCDPA), S. 3663, introduced by Senators Roger Wicker, John Thune, Jerry Moran, Marsha Blackburn, and Deb Fischer on May 7, 2020;
2. the Public Health Emergency Privacy Act (PHEPA), companion bills S. 3749 and H.R. 6866, introduced, respectively, by Senators Richard Blumenthal and Mark Warner and Representatives Anna Eshoo, Janice Schakowsky, Suzan DelBene, Yvette Clarke, G.K. Butterfield, and Tony Cardenas on May 14, 2020; and

[9] Source: https://ec.europa.eu/info/live-work-travel-eu/health/coronavirus-response/digital-solutions_en#european-supercomputers-fighting-the-coronavirus

[10] Source: https://ec.europa.eu/digital-single-market/en/news/coronavirus-eu-interoperability-gateway-goes-live-first-contact-tracing-and-warning-apps-linked

3. the Exposure Notification Privacy Act (ENPA), S. 3861, introduced by Senators Maria Cantwell and Bill Cassidy on June 1, 2020.

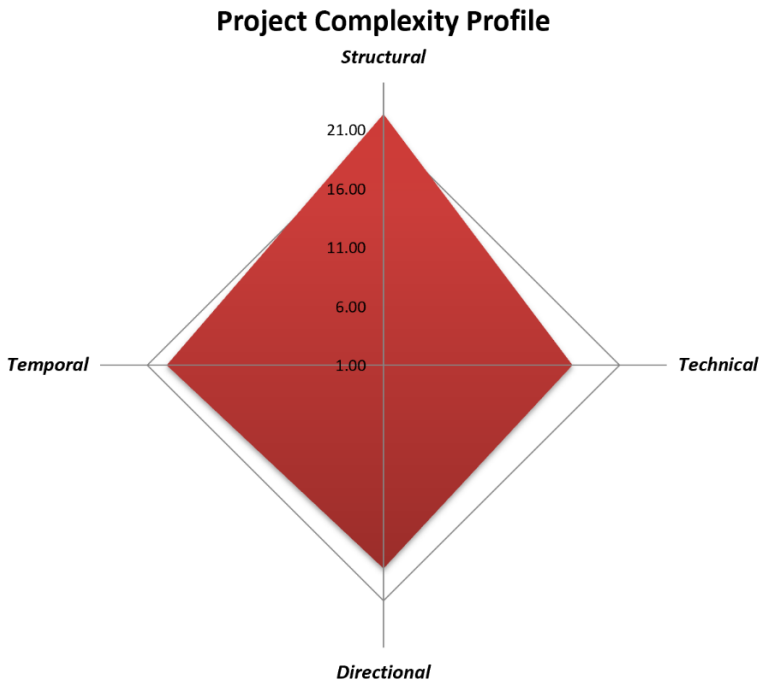
Those three bills take a similar approach to regulate contact-tracing data. e.g., a covered entity would have to take certain procedures before and after collecting covered data. In general, those proposed bills will expand the definitions or the scopes of the covered entity and covered data. For instance, the covered entity will include any entity or persona engaged in contact-tracing and exposure notification or develop tools for contact-tracing and exposure notification. Moreover, each bill will protect the specific data collected (such as geolocation data or any information linked or reasonably linked to any individual), besides personal health information.

The system perspective notes that interactions among the general public, the public health authorities, and the regulators create challenges such that a trade-off between the cost of pandemic and data privacy results. The apparent conflict is that the fast spread of COVID-19 leaves little time for legal and ethical debates of personal data privacy and remedies of a complicated U.S. health data privacy system. Remington and Pollack (2007) suggest four types of project complexity i.e., structural, technical, directional, and temporal complexity that may describe how to understand the adoption of COVID-19 tracing apps in the U.S. (see Table 5). The structural, directional, and temporal complexity are all ranked high because of the lack of a grand strategy, coordination of federal and state government, and collective efforts. In contrast, the technical complexity is rated as medium not because we do not have a disposable technology but because the related legal and ethical issues of data privacy complicate the choice and use of technologies. Figure 5 provides the project complexity profile for digital contact-tracing.

Table 5: Mapping Project Complexity for Digital Contact-Tracing in U.S.

Dimension	Evaluation	Project Complexity
Structural (number of dependencies)	• The fragmented state-by-state approach in developing and adopting COVID-19 tracing apps in the US greatly limits the effectiveness of such a tactic (Barber & Knight 2020, Timberg et al., 2020). • States have fewer incentives to invest in COVID-19 because the federal government does not provide them any financial supports • Digital contact-tracing must rely on other successful measures, including fast and cheap testing, quick follow-up, and effective quarantine and isolation procedures.	High
Technical (Impact of unresolved technical/design issues	• Bluetooth-based contact -tracing apps have some limitations. For instance, the current Apple-Google protocol emphasizes personal data privacy, and the apps do not collect key personal information such as location data, which is essential to containing the spread of the COVID-19 from the public health perspective. • Legal challenges related to personal data privacy in the U.S., i.e., how to protect the information and entities involved in digital contact-tracing.	Medium
Directional (ambiguity/lack of agreement on goals)	• Fighting a pandemic like COVID-19 requires a clear vision and strong leadership, quick responses based on scientific evidence, and collective efforts from the individuals and the society as a whole. However, there is no clear direction regarding how and what to do to fight the pandemic. • The general public's trust and cooperation in the effectiveness of digital contact-tracing and other public health measures (mask, social distancing, etc.)	High
Temporal (Expected time delays at key project stages)	• The value and importance of digital contact-tracing could change over time. For instance, the priority for contact-tracing could be lower for states with high positive testing rates that combat the congested health care system. Moreover, the advent of the COVID-19 vaccine could also compete for the limited financial resources at both the federal and state level.	High

Figure 5: Project Complexity Profile for Digital Contact-Tracing in U.S.



Big data analytics, along with the advanced analytical tools such as AI and ML, has the potential to change the landscape of the health care systems, for example, by optimizing workflows in healthcare providers, providing more accurate diagnoses, and improving the overall quality of services to patients and the general public. However, it also raises challenges for the regulators worldwide regarding how to minimize the potential "harms," such as data privacy, brought by these new analytical tools and decision-making models.

To do so, certain essential features of complex systems in health care must be considered. Unlike data systems, human systems are purposeful, self-organizing and constantly adapt to change; they are driven by the interactions between systems components and governed by meaningful feedback; and they are nonlinear and hard or impossible to predict, with changes in one part of the system causing unexpected changes in other sub-systems. The development of big data analytics and its applications in various business decision-making processes share some similar features. For instance, the analytical tools are fast evolving, and the ways that data are collected, shared, and analyzed are fast changing. Therefore, a "systems" perspective is vitally important (Gerke et al., 2020) and argues for consideration of the following recommendations for effective regulation of big data analytics in health care:

1. Due to the large number of elements that interact in nonlinear and dynamics ways, regulators should take a holistic view because the entire value chain of big data analytics

can go beyond the traditionally defined health care industry with the participation of stakeholders outside the current laws and regulations.

2. Due to individual interactions, influences and relationships, organizational should embark on a continuous learning process to discover emerging patterns as a foundation for a more effective regulatory approach. Because big data analytics is fast evolving, and the context where big data analytics are applied is rapidly changing, there is no clear separation between the "benefit" and "harm." Regulators should understand the context and this should be reflected in the history of the system. In particular, the risk assessment of the application of AI/ML must reflect the changes in external conditions and the system itself.
3. All stakeholders should realize that the people in the system shape the system and are influenced by the system. Regulators should emphasize a human-centric regulatory approach and develop policies that will foster positive feedback (such as trust, cooperation) from the people in the system.

Xuejuan Zhang is a candidate for the degree of Doctor of Management in Strategic Leadership in the School of Continuing and Professional Studies at Thomas Jefferson University. This paper was submitted as an advanced independent study to Larry M. Starr, PhD. Ms. Zhang can be reached at Xuejuan.Joyce.Zhang@Jefferson.edu.

References

Agrawal, R. & Prabakaran, S. 2020. Big data in digital healthcare: lessons learned and recommendations for general practice. *Heredity*, 124: 525-534.

Bleicher, A. 2017. Demystifying the Black Box That is AI. *Scientific American*, August 9. Retrieved from: https://www.scientificamerican.com/article/demystifying-the-black-box-that-is-ai/

Colson, E. 2019. What AI-Driven Decision Making Looks Like. *Harvard Business Review*, July 8. Retrieved from: https://hbr.org/2019/07/what-ai-driven-decision-making-looks-like

Davis, J. 2019. How the federal data privacy debate, regulations may impact healthcare. *HealthITSecurity,* part of the *Xtelligent Healthcare Media network*, March 4. Retrieved from: https://healthitsecurity.com/news/how-the-federal-data-privacy-debate-regulations-may-impact-healthcare

Evans, R. S. 2016. Electronic Health Records: Then, Now, and in the Future. *IMIA Yearbook of Medical Informatics*, Supp. 1: S48-61. Retrieved from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5171496/

European Commission. 2020. *On Artificial Intelligence – A European Approach to Excellence and Trust*, White Paper, COM/2020/65 final. Retrieved from: https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

Executive Office of the President. 2014. Big Data: Seizing Opportunities, Preserving Values. *White House*. Retrieved from: https://obamawhitehouse.archives.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf

Figliola, P. M. 2020. Digital Contact Tracing Technology: Overview and Considerations for Implementation. *Congressional Research Service*, IF11559. Retrieved from: https://fas.org/sgp/crs/misc/IF11559.pdf

Gaffney, J. M. 2020. Tracing Papers: A Comparison of COVID -19 Data Privacy Bills. *Congressional Research Service Legal Sidebar*, LSB 10501. Retrieved from: https://crsreports.congress.gov/product/pdf/LSB/LSB10501

Gerke, S., T. Minssen, & Cohen, I. G. 2020. Ethical and legal challenges of artificial intelligence-driven healthcare. In *Artificial Intelligence in Healthcare* ed by Adam Bohr and Kaveh Memarzadeh. Academic Press. Retrieved from: https://www.sciencedirect.com/science/article/pii/B9780128184387000125

Gerke, S., B. Babic, T. Evgeniou & Cohen, I. G. 2020. The need for a system view to regulate artificial intelligence/machine learning-based software as medical device. *NPI Digital Medicine* 3, Article number: 53. Retrieved from: https://www.nature.com/articles/s41746-020-0262-2#citeas

Grady, N. 2019. NIST Big Data Interoperability Framework: Volume 1, Definitions. *National Institute of Standard and Technology (NIST),* Special Publication (NIST SP) - 1500-1r2, Version 3. Gaithersburg, MD: U.S. Department of Commerce. Retrieved from: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-1r2.pdf

Greenberg, P. 2019. States Break New Ground on Consumer Privacy Regulation. *NCLS*. Retrieved from: https://www.ncsl.org/blog/2019/06/19/states-break-new-ground-on-consumer-privacy-regulation.aspx#:~:text=Nevada%20and%20Maine%2C%20however%2C%20enacted,with%20their%20legislation%20this%20year.&text=Now%2C%20operators%20of%20Internet%20websites,request%20it%20not%20be%20sold.

Health IT Policy Committee (HITPC) Privacy and Security Workgroup. 2015. Health big data recommendations. *Office of the National Coordinator for Health Information Technology, (ONC)*. Retrieved from: https://www.healthit.gov/sites/default/files/facas/HITPC_Health_Big_Data_Report_FINAL.pdf

Mulligan, S. P., C. D. Linebaugh, & Freeman, W.C. 2019. Data Protection Law: An Overview. Congressional Research Service Report R45631. Retrieved from: https://crsreports.congress.gov/product/pdf/R/R45631.

National Conference of State Legislatures (NCSL). 2019. Cybersecurity Legislation 2019. Retrieved from: https://www.ncsl.org/research/telecommunications-and-information-technology/cybersecurity-legislation-2019.aspx.

OECD. 2013. OECD guidelines on the protection of privacy and transborder flows of personal data. Retrieved from: https://www.oecd.org/internet/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm

President's Council of Advisors on Science and Technology. 2014. Big Data and Privacy: A Technological Perspective. Retrieved from: https://bigdatawg.nist.gov/pdf/pcast_big_data_and_privacy_-_may_2014.pdf

Shah, A., & Hedgeman, B. 2020. Public Health vs. Patient Privacy – How Coronavirus is Putting HIPPA to the Test. National Law Review, Volume X, Number 313. Retrieved from: https://www.healthlawadvisor.com/2020/03/03/public-health-vs-patient-privacy-how-coronavirus-is-putting-hipaa-to-the-test/

Terry, N. P. 2017. Regulatory disruption and arbitrage in healthcare data protection. Yale Journal of Health Policy, Law, and Ethics 17(1): 143-208. Retrieved from: https://digitalcommons.law.yale.edu/yjhple/vol17/iss1/3/

Thorpe, J. H., & Gray, E. A. 2015. Big Data and Public Health: Navigating Privacy Laws to Maximize Potential. Public Health Reports, 130(2): 171-175. Retrieved from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4315864/

Timberg, C., Hendrix, S., Kim, M. J. & Weber-Steinhaus, F. 2020. Cellphone apps designed to track COVID-19 spread struggle worldwide amid privacy concerns. Washington Post, August 18, 2020. Retrieved from: https://www.washingtonpost.com/technology/2020/08/17/covid-tracking-apps-cellphones/