International Telecommunication Union

FINANCIAL INCLUSION GLOBAL INITIATIVE (FIGI)

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

(12/2018)

Security, Infrastructure and Trust Working Group

Big data, machine learning, consumer protection and privacy

Report of Trust Workstream

Security, Infrastructure and Trust Working Group: *Big data, machine learning, consumer protection and privacy*

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, information and communication technologies (ICTs). The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

A new global program to advance research in digital finance and accelerate digital financial inclusion in developing countries, the Financial Inclusion Global Initiative (FIGI), was launched by the World Bank Group, the International Telecommunication Union (ITU) and the Committee on Payments and Market Infrastructures (CPMI), with support from the Bill & Melinda Gates Foundation.

The Security, Infrastructure and Trust Working Group is one of the three working groups which has been established under FIGI and is led by the ITU. The other two working groups are the Digital Identity and Electronic Payments Acceptance Working Groups and are led by the World Bank Group.

Big data, machine learning, consumer protection and privacy

Trust Workstream

Security, Infrastructure and Trust Working Group: *Big data, machine learning, consumer protection and privacy*

About this Paper

This report has been written by Rory Macmillan, Partner, Macmillan Keck Attorneys & Solicitors with input and comment from […]

Security, Infrastructure and Trust Working Group: *Big data, machine learning, consumer protection and privacy*

Table of contents

Contents

1 Executive Summary

This paper explores various challenges that consumer protection and data privacy law and regulation face with regard to big data and machine learning techniques, particularly where these are used for making decisions about services provided to consumers.

The beneficial opportunity data presents for development is widely recognised, particularly for the provision of digital financial services. Service providers can use big data to build a detailed personal profile of an individual including his or her behaviour (e.g., preferences, activities and movements) which may be used for commercial offers. Big data and machine learning are being increasingly deployed for financial inclusion, not only in wealthy nations but also in developing countries. These new technologies also bring risks, some say tendencies, of bias in decision-making, discrimination and invasion of privacy.

Artificial intelligence involves techniques that seek to approximate aspects of human or animal cognition using computing machines. Machine learning refers to the ability of a system to improve its performance, by recognising patterns in large datasets. Big data relies upon and is typically defined by, computer processing involving high volumes and varieties of types of linked up data processed at high velocity (the "three Vs" – sometimes expanded to four Vs by the addition of "veracity").

Consumer protection involves the intervention of the State through laws and processes in what would otherwise be a private relationship between consumer and provider. It aims to compensate for perceived information, bargaining and resource asymmetries between providers and consumers.

Increasingly, countries are legislating to protect the personal data and privacy of their subjects, granting them rights that give them more power over how their personal data is used. These laws are under strain in an era of big data and machine learning. Complying with requirements to notify the consumer as to the purpose of data collection is difficult where, as in machine learning, the purpose may not be known at time of notification. Consent is difficult to obtain when the complexity of big data and machine learning systems is beyond the consumer's comprehension. The notion of data minimisation (collecting and storing only data necessary for the purpose for which it was collected, storing it for the minimum period of time) runs counter to the modus operandi of the industry, which emphasises maximizing the volumes of data collection over time. As stated in a 2014 report to the US President in 2014, "The notice and consent is defeated by exactly the positive benefits that big data enables: new, non-obvious, unexpectedly powerful uses of data."

Some suggest privacy expectations are highly contextual. Tighter restrictions on collection, use and sharing of personal data in some situations (and tiered consent which differentiates between types of data according to use or the organisation that may use it) have been discussed. Sunset clauses providing that the individual's consent to use his or her personal data will expire after a period of time (and potentially renewed) have also been suggested. Efforts are also being made to develop technologies and services to manage consent better. There appears to be a genuine commercial opportunity for investment and innovation to improve management of such consumer consent.

The successful functioning of machine learning models and the accuracy of their outputs depends on the quality of the input data. Data protection and privacy laws increasingly impose legal responsibility on firms to ensure the accuracy of the data they hold and process. However, they do not legislate for accuracy of output from big data and machine learning systems. This raises questions about the regulatory responsibilities of those handling big data, concerning both the accuracy of input data in

automated decisions and the data reported in formal credit data reporting systems. In some jurisdictions, this has given rise, among other remedies, to certain rights to object to automated decisions.

Inferences from input data generated by machine learning models determine how individuals are viewed and evaluated for automated decisions. Data protection and privacy laws may be insufficient to deal with the outputs of machine learning models that process such data. One of their concerns is to prevent discrimination, typically protecting special categories of groups (e.g., race, ethnicity, religion, gender). In the era of big data, however, non-sensitive data can be used to infer sensitive data.

Machine learning may lead to discriminatory results where the algorithms' training relies on historical examples that reflect past discrimination, or the model fails to consider a wide enough set of factors. Addressing bias is challenging, but tests have been developed to assess where it may arise. In some countries, where bias is unintentional, it may nevertheless be unlawful if it has "disparate impact," which arises where the outcomes from a selection process are widely different for a protected class of persons.

A key question is to what degree firms should bear the cost of identifying potential bias and discrimination within their data algorithms. Firms relying on big data and machine learning might employ tools (and under some laws be responsible) to ensure that their data will not amplify historical bias, and to use data to identify discrimination. Ethical frameworks and "best practices" may be needed to ensure that outcomes will be monitored and evaluated, and algorithms adjusted.

The vast amounts of data held by and transferred among big data players creates risks of data security breach, and thus risk to consumer privacy. Personal privacy may be protected in varying degrees by using privacy enhancing technologies (PETs). A market is growing in services for de-identification, pseudonymisation and anonymisation. Differential privacy is also increasingly being employed. Regulation may need to ensure that privacy enhancing technologies are continuously integrated into big data and machine learning data processing. This may require establishing incentives in legislation that create liability for data breaches, essentially placing the economic burden not on the consumer by obtaining their consent but on the organisations collecting, using and sharing the data.

Big data and machine learning are made possible by intermediaries, such as third-party data brokers who trade in personal data. Transfer of personal data creates risk of breach and identity theft, intrusive marketing and other privacy violations. Data brokers are coming under increasing scrutiny, and laws providing consumers direct rights are being introduced.

Conventional requirements to provide notice of the intended purpose of using a consumer's personal data when the purpose may as yet be unclear, or obtaining consent for something the consumer largely cannot understand, are under strain. Risks from inaccuracy of data inputs, or bias and discriminatory treatment in machine learning decisions also raise difficult questions about how to ensure that consumers are not unfairly treated. The difficulty of ensuring transparency over decisions generated by algorithms, or of showing what harm has been caused by artificial intelligence techniques that would not have otherwise been caused, also pose challenges for consumer protection and data privacy law and regulation.

The challenges arising for the treatment of big data and machine learning under legal and regulatory frameworks for data protection and privacy suggest that the development of robust self-regulatory and ethical regimes in the artificial intelligence and financial services community may be particularly

important. Facing legal and regulatory uncertainty, businesses may introduce risk management systems, employ privacy by design and develop ethics.

There are various areas for further exploration and development of standards and procedures, including in relation to acceptable inferential analytics, reliability of inferences, ethical standards for artificial intelligence, provision of post-decision counterfactuals, documentation of written policies, privacy principles for design, explanations of automated decisions, access to human intervention, and other accountability mechanisms.

Security, Infrastructure and Trust Working Group: *Big data, machine learning, consumer protection and privacy*

2 Acronyms

CRISP/DM	Cross-industry process for data mining
DPIA	Data Protection Impact Assessment
FAT	Fairness, accountability and transparency
FCRA	US Fair Credit Reporting Act
FTC	US Fair Trade Commission
GDPR	EU General Data Protection Regulation
GPCR	World Bank's General Principles on Credit Reporting
KYC	Know your customer
NIST	National Institute of Standards and Technology
PET	Privacy enhancing technology

3 Introduction

Big data, artificial intelligence and machine learning are dominating the public discourse, whether from excitement at new capabilities or fears of lost jobs and biased automated decisions. The issues are not entirely new.[1] However, public awareness of the potential of powerful computing systems applying complex algorithms to huge volumes of data has grown with stories of computers beating humans at games and as people increasingly enjoy services produced by such systems.[2]

Personal identifiable data is widely collected, shared and available on commercial data markets. Such data may include an individual's internet and transaction history, registration with public and private organisations, and use of social media. Firms and governments routinely collect, process and share such data with third parties, often without the user's knowledge or consent.

The beneficial opportunity data presents for development is widely recognised, particularly for the provision of digital financial services.[3] Many financial services depend on risk assessment and management. For example, a loan's value is in large part based on the borrower's creditworthiness, as well as the collateral that may secure the loan. The more data there is about the borrower, the better the lender can assess their creditworthiness. Big data enables inferences about creditworthiness to be drawn from a borrower's membership of one or more categories of persons who have borrowed and repaid or defaulted on debts in the past.

Digital financial service providers can not only generate commercial profit but, with information about and analysis of consumers' background and interests, can also add substantial public value through improved access to financial services.

Artificial intelligence is increasingly used to analyse a wide range of data sources to create a coherent assessment of consumers' creditworthiness and make lending decisions. Instead of relying merely on the borrower's representation of income and existing debts in the loan application, or an interview by the local bank manager, or checking a credit reporting agency's score (e.g., FICO), the combination of artificial intelligence and big data allows firms to analyse an individual's digital footprint to predict the probability of default. This enables access to services that may otherwise have been unavailable.

Big data analytics may be used to enhance traditional means of credit assessments. Credit reference bureaus, such as Equifax, have claimed to have made significant improvements in the predictive ability of their models by using big data analytics. This can be particularly useful in assessing individuals who lack a traditional credit history, thus giving them access to credit services. This opportunity extends beyond enhancing traditional means of credit assessment to entirely new models. For example, Upstart[4] uses machine learning to predict young adults' creditworthiness drawing from data on their education, exam scores, academic field of study and job history data, in an automated loan process. It offers loans directly to consumers, as well as offering other lenders its software as a service, i.e., a platform for their own lending services.

---

[1] The risk of bias in computing systems and approaches to dealing with it have been under discussion for more than 20 years. See Batya Friedman & Helen Nissenbaum, *Bias in Computer Systems*, 14 ACM TRANSACTIONS ON INFO. SYS. 330 (1996).

[2] *See, e.g.*, Cade Metz, *In a Huge Breakthrough, Google's AI Beats a Top Player at the Game of Go*, WIRED (Jan. 27, 2016), https://www.wired.com/2016/01/in-a-huge-breakthrough-googles-ai-beats-a-top-player-at-the-game-of-go/.

[3] *A World That Counts: Mobilizing the Data Revolution for Sustainable Development*, United Nations Secretary-General's Independent Expert Advisory Group on a Data Revolution for Sustainable Development.

[4] https://www.upstart.com/.

These business models are being increasingly deployed for financial inclusion not only in wealthy nations but also in developing countries. Lenndo,[5] a fintech firm supporting credit evaluation with alternative data analysis, has partnered with the global credit agency FICO to make FICO score services available in India.[6] This service evaluates alternative data from a consumer's digital footprint to produce a credit score for those who do not have sufficient traditional data on file ("thin file" borrowers) with one of the Indian credit bureaus for a traditional loan approval. Branch.co[7] and MyBucks[8] are active in Africa and beyond, using identity proofing and automated mobile app that uses credit-scoring engines to generate credit scores from analysing a customer's mobile phone bill, text messages, payment history, bank account history (if the person has a bank account), utility bills and geolocation data.

Rapid access to large volumes of data is key to the effectiveness of such technologies. For instance, ZestFinance has a strategic agreement with its investor Baidu, the Chinese internet search provider (equivalent of Google in China) that allows ZestFinance to access individuals' search history, geolocation and payment data to build credit scores in China, where around half of the population has no credit history.[9] ZestFinance's CEO famously said, "all data is credit data."[10]

Artificial intelligence is not only useful for credit risk assessment. Any service involving risk assessment depends on information and analysis. The firm Progressive, for example, collects data on individuals' drivers' driving performance through mobile applications like Snapshot in order to predict risk of accidents and offer (or not) discounted insurance premiums.[11] Artificial intelligence is being used in numerous other applications in the field of insurance.[12] Other areas where artificial intelligence is having a substantial impact on innovation and improvements to efficiency include personalisation of savings products, management of payment services, provision of virtual assistance for customers (e.g., robo-advisory and chatbots), and detection of fraud, money laundering and terrorism financing.

The rise in consumer use of products and services relying on artificial intelligence and machine learning has triggered a vigorous policy debate about its risks, and the need for coherent policy.[13] The IEEE's Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems, for example, has called for legislators to consider regulation:[14]

> *Lawmakers on national, and in particular on international, levels should be encouraged to consider and carefully review a potential need to introduce new regulation where appropriate, including rules subjecting the market launch of new AI/AS driven technology to prior testing and approval by appropriate national and/or international agencies.*

[5] https://www.lenddo.com/.
[6] http://www.prnewswire.co.in/news-releases/new-fico-credit-scores-provide-lenders-opportunity-to-expand-access-to-credit-in-india-for-nearly-350-million-653029163.html.
[7] https://branch.co/.
[8] https://corporate.mybucks.com/.
[9] https://www.businesswire.com/news/home/20160717005040/en/ZestFinance-Receives-Funding-Baidu-Fuel-Development-Search-Based.
[10] https://www.pymnts.com/in-depth/2015/how-zestfinance-used-big-data-lending-to-secure-150m-from-fortress/.
[11] https://www.progressive.com/auto/discounts/snapshot/.
[12] See https://www.techemergence.com/machine-learning-at-insurance-companies/.
[13] See, e.g., Ryan Calo, Artificial Intelligence Policy: A Primer and Roadmap, https://ssrn.com/abstract=3015350.
[14] See Ethically Aligned Design, at footnote 217.

Longstanding laws and regulations that aim to protect consumers from adverse uses of personal data are facing various challenges in terms of new data collection and analytical methodologies. Indeed, some have ventured to say that even the most recent of data protection and privacy laws, Europe's General Data Protection Regulation (GDPR), sometimes referred to as the "gold standard" of data protection and privacy law, is "incompatible" with the world of big data.[15] Similar concerns arise in relation to other global standards, such as the OECD Recommendation Concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data (the OECD Privacy Guidelines)[16] and the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (ETS No. 108) (referred to as Convention 108), as recently amended by the Amending Protocol to the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data.[17]

Three core tenets of data protection and privacy law are purpose specification, data minimisation, and the treatment of data of "protected" or "special" categories of groups (such as racial, gender, religious and other groups). These tenets come under strain when the specific purpose of collecting and processing data may only become understood as the machines themselves learn from high volumes of observed and performance data, producing more accurate analysis. Personal data can also serve as a proxy for membership of a protected group.

These new technologies also present risks, some even say tendencies, of bias in decision-making, discrimination and invasion of privacy.[18] Analytics may be used to draw inferences (and in some cases make predictions) about a person's race, gender, sexual orientation, relationships, political views, health (including specific disease), mental state, personal interests, creditworthiness and other attributes. Discrimination may be embedded in the data processing, effectively leading to results that would be prohibited by gender or race discrimination laws if decisions were carried out through human (as opposed to machine) processes.

These risks are particularly relevant to financial services. Unlike many consumer products and services, offers and pricing of financial services depend on the profile of the individual consumer. The decision to offer a loan, and at what interest rate, the decision to issue a credit card, and with what credit limit, and the decision to offer different types of insurance, all depend on assessing the risk the individual presents. Thus, like the decision to employ or not to employ someone, many financial services have an important personal dimension.[19]

---

[15] "The GDPR's provisions are—to borrow a key term used throughout EU data protection regulation—incompatible with the data environment that the availability of Big Data generates. Such incompatibility is destined to render many of the GDPR's provisions quickly irrelevant." Zarsky, Tal, 'Incompatible: The GDPR in the Age of Big Data' (August 8, 2017). Seton Hall Law Review, Vol. 47, No. 4(2), 2017, https://ssrn.com/abstract=3022646.

[16] See https://www.oecd.org/sti/ieconomy/2013-oecd-privacy-guidelines.pdf.

[17] The Amending Protocol to the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, which amended Convention 108 in 2018, now addresses features of automated data processing such as profiling, automated decisions and use of algorithms. This includes the right not to be subject to a decision significantly affecting a data subject based solely on automated data processing without considering their view; the right to obtain knowledge of the reasoning underlying data processing where the results are applied to the data subject; and the right to object to data processing, among others.

[18] Danielle Citron and Frank A Pasquale, 'The Scored Society: Due Process for Automated Predictions' (Social Science Research Network 2014) SSRN Scholarly Paper ID 2376209 https://papers.ssrn.com/abstract=2376209; Tal Z Zarsky,'Understanding Discrimination in the Scored Society' (2014) 89 Wash. L. Rev. 1375; Brent Mittelstadt and others, 'The Ethics of Algorithms: Mapping the Debate' (2016) 3 Big Data & Society http://bds.sagepub.com/lookup/doi/10.1177/2053951716679679.

[19] This is not so for all financial services; for instance, the retail deposit business of a bank, or an investment fund available for retail investors, have no particular reason to treat investors differently.

This can enable services to be better tailored to the individual's risk profile, and thus facilitates access to financial services that might otherwise not have been offered. However, at the same time, the individual may be unaware of the data relied on to draw inferences or the reason for a decision not to extend services to them, and may lack a way to dispute the data, inferences and decision.

Access to data about individuals enables such decisions to be based increasingly on individual behaviour, but with potential invasion of privacy. In 2008, the US Federal Trade Commission intervened to stop unfair practices by CompuCredit, which marketed credit cards to people with subprime credit. CompuCredit had been reducing consumers' credit limits based on a model that reduced their scores where they engaged in certain transactions, such as visiting pawn shops, personal counselling and pool halls.[20]

The treatment of data available on individuals, and in particular the process of profiling them and drawing inferences about them, is thus central to the provision of such financial services. Consequently, achieving fairness, accuracy and transparency in financial services must take into account what and how personal data is being collected, being used, and being shared with third parties.[21]

These challenges are made more complex by the variety of regulatory frameworks applying to different types of digital financial service providers, some of which are regulated as banks, and others of which are barely regulated at all. Even when they provide similar services, different restrictions may apply to the data they may collect and use, and different remedies may be available for consumers.

The challenges arising for the treatment of big data and machine learning under legal and regulatory frameworks for data protection and privacy suggest that the development of robust self-regulatory and ethical regimes in the artificial intelligence and financial services community may be particularly important.

This paper provides background for policy makers, regulators, digital financial service providers, investors and other organisations concerning the need for solutions and standards on protecting consumer data privacy in the context of big data and machine learning. These issues are still emerging as the technologies, use cases and adoption rapidly increase. As a result, while the issues are increasingly understood, there are few areas in which there is widespread consensus on definitive best practices. Approaches will depend on how policy makers, legislators, regulators and market participants weigh up trade-offs and synergies among policy objectives such as experimentation and innovation, economic productivity, trust in services, and consumer protection.

This paper explores various views, citing organisations', academics', and thinkers' suggestions on commonly adopted approaches to protecting consumer data privacy and the associated laws and regulations. The purpose of this paper is to highlight these ideas and not to take a position. It seeks to

[20] Ryan Singel, *Credit Card Firm Cut Limits After Massage Parlor Visits, Feds Allege*, Wired, 20 June 2008, https://www.wired.com/2008/06/credit-card-fir/; FTC complaint at https://www.wired.com/images_blogs/threatlevel/files/compucreditcmplt.pdf; Subprime Credit Card Marketer to Provide At Least $114 Million in Consumer Redress to Settle FTC Charges of Deceptive Conduct, 19 December 2008, https://www.ftc.gov/news- events/press-releases/2008/12/subprime-credit-card-marketer-provide-least-114-million-consumer; and FTC v CompuCredit Corporation and Jefferson Capital Systems LLC, Civil No. 1:08-CV-1976-BBM-RGV, Stipulated Order for Permanent Injunction and Other Equitable Relief Against Defendant CompuCredit Corporation, https://www.ftc.gov/sites/default/files/documents/cases/2008/12/081219compucreditstiporder.pdf.

[21] See generally, World Bank, *New Forms of Data Processing Beyond Credit Reporting: Consumer and Privacy Aspects*, 2018; and Responsible Finance Forum, *Opportunities and Risks in Digital Financial Services: Protecting Consumer Data and Privacy*, 2017.

support those who must wrestle with these matters at a policy, legislative and regulatory level in the coming years. Rather than recommending best practices, this paper therefore focuses on identifying and framing key issues for consideration when developing regulatory frameworks (including potentially self-regulatory frameworks).[22]

Section 4 introduces the key concepts in play, starting with the technology and market trends of big data and machine learning (section 4.1), and then the regulatory dimensions of consumer protection (section 4.2) and data privacy (section 4.3).

The paper then proceeds to consider consumer protection and data privacy in three broad phases of the consumer's encounter with service providers that rely on big data and machine learning:

Section 5 discusses the pre-engagement phase, which primarily concerns what disclosures and notifications are required to be made to consumers about how and for what purpose their personal data will be collected, used and transferred to third parties, and requirements for obtaining consumer consent to legitimise use of personal data.

Section 6 discusses the engagement phase, which relates to the restrictions on, requirements relating to, and responsibility for the things firms may do with personal data, including in relation to accuracy in machine learning models (section 6.1), bias and discriminatory treatment (section 6.2), data breach and re-identification (section 6.3), and transfer of data to third parties (section 6.4).

Section 7 turns to the post-engagement phase, and the consumer's means of holding big data and machine learning operators accountable for violations of consumer protection and data privacy laws. It looks at consumers' rights to access personal data about themselves, rectifying errors in it and requesting that it be erased (section 7.1), transparency difficulties with obtaining explanations for complex machine learning model outputs (section 7.2), the right to contest decisions and obtain human intervention (section 7.3), and the challenge of showing harm (section 7.4).

The paper discusses in section 8 some practical steps firms may take to reduce risk in face of the legal and regulatory uncertainties. It closes in section 0 with a short list of areas for further development in this field, whether in the development of ethics, standards or procedures.

[22] This paper does not cover all data privacy issues, or all consumer protection issues that arise in relation to personal data. Nor does this paper cover all aspects of big data and machine learning. Many outputs of these techniques are general to society and are useful for health, education and other policies, but do not have a direct impact through decisions made about specific individuals. As a result, some rights and obligations are explored in more detail than others, focusing on where big data and machine learning pose particular challenges to data privacy and consumer protection.

4 The scope of discussion

4.1 What are big data and machine learning?

Artificial intelligence involves techniques that seek to approximate aspects of human or animal cognition using computers. ***Machine learning***, a form of artificial intelligence, refers to the ability of a system to improve its performance, often by recognising patterns in large datasets, doing so at multiple layers of analysis (often referred to as deep learning).[23]

Machine learning algorithms build a model from ***training data***, i.e., historical examples, in order to make predictions or decisions rather than following only pre-programmed logic. Neural networks analyse data through many layers of hardware and software.[24] Each layer produces its own representation of the data and shares what it "learned" with the next layer. Machine learning learns by example, using the training data to train the model to behave in a certain way.[25] Machine learning is not new, but as a result of big data, it is suddenly being deployed in numerous practical ways.[26]

Figure 1 Machine learning, xkcd.com

Big data relies upon and is typically defined by, computer processing involving high volumes and varieties of types of linked up data processed at high velocity (the "three Vs"[27] – sometimes expanded to four Vs by the addition of "veracity").[28] The advent of big data techniques arises from developments in how data is collected, stored and used. Data is collected using numerous applications and sensors which record consumers' communications, transactions and movements. Distributed databases store the data, and high-speed communications transmit it at high speed, reducing the cost of data analytics. Advanced analytical processes are applied in numerous contexts.

In the financial services context, big data may include ***alternative data***, i.e., data that is not collected and documented pursuant to traditional credit reporting, but rather collected from a wide range of other

[23] Harry Surden, *Machine Learning and the Law*, 89 WASH. L. REV. 87, 88 (2014).
[24] F. Rosenblatt, Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain, Psychological Review, Vol 65, No. 6, 1958 http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.335.3398&rep=rep1&type=pdf
[25] Commonly known examples are IBM Watson, Google/Deepmind Alphago, Apple Siri and Amazon Alexa, all of which rely on machine learning to advance their service for the user.
[26] Peter Stone et al., Stanford Univ., Artificial Intelligence and Life in 2030: Report of the 2015-2016 Study Panel 50 (2016), https://ai100.stanford.edu/sites/default/files/ai_100_report_0831fnl.pdf.
[27] Doug Laney, 3D Data Management: Controlling Data Volume, Velocity and Variety, Metra Group Research Note (2001) 6.
[28] IBM, The Four V's of Big Data (2014), http://www.ibmbigdatahub.com/infographic/four-vs-big-data; Viktor Mayer-Schönberger and Kenneth Cukier, *Big Data: A Revolution That Will Transform How We Live, Work and Think* (John Murray 2013).

digital sources, such as social media and electronic transaction history.[29] Large quantities of data from use of computer browsers and mobile phone apps may be collected and shared without being subject to standard opt-in policies. For instance, a recent Oxford University study of about 1 million Android apps found that nearly 90 per cent of apps on Android smartphones transfer information to Google.[30] Customer internet usage may be swept up along with location data, contact information and text messages (see Figure 2).

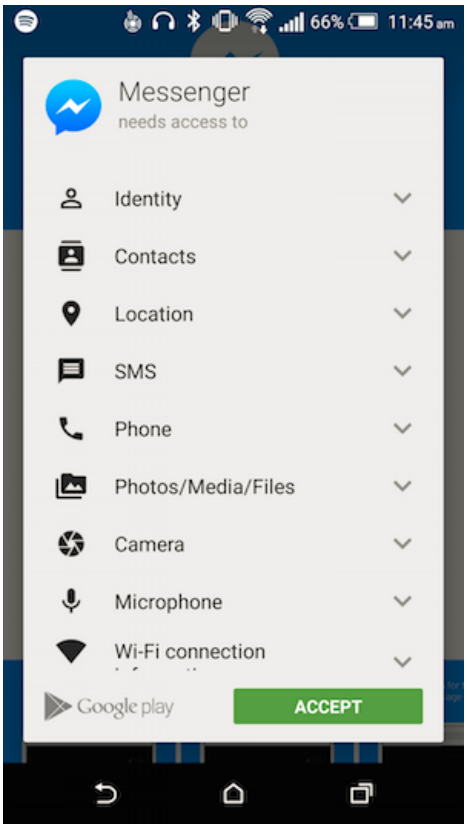

Figure 2. Smartphone app permission settings

One adviser to investors in big data market players lists the following sources of alternative data available in today's market:[31]

- Data from financial aggregators
- Credit card data

[29] GPFI, Use of Alternative Data to Enhance Credit Reporting to Enable Access to Digital Financial Services by Individuals and SMEs operating in the Informal Economy, Guidance Note, PREPARED BY INTERNATIONAL COMMITTEE ON CREDIT REPORTING (ICCR) (GPFI Priorities Paper 2018). https://www.g20.org/sites/default/files/documentos_producidos/use_of_alternative_data_to_enhance_credit_reporting_to_enable_access_to_digital_financial_services_iccr.pdf

[30] Reuben Binns, Ulrik Lyngs, Max Van Kleek, Jun Zhao, Timothy Libert, Nigel Shadbolt, *Third Party Tracking in the Mobile Ecosystem*, arXiv:1804.03603v3 [cs.CY] 18 Oct 2018, https://arxiv.org/pdf/1804.03603.pdf; Aliya Ram, Aleksandra Wisniewska, Joanna S. Kao, Ændrew Rininsland, Caroline Nevitt, *How smartphone apps track users and share data*, Financial Times, 23 October 2018, https://ig.ft.com/mobile-app-data-trackers/.

[31] ZwillGen, Alternative Data: Best Practices, presented at the Privacy and Security Forum in Washington DC, 2018.

- Geospatial and location data
- Web scraping datasets
- App engagement data
- Shipping data from U.S. customs
- Ad spend data
- Data made available through APIs
- Location/foot traffic data from sensors and routers
- Social media data
- B2B data acquired from parties in the supply chain
- Agriculture data (e.g., feeds on corn production)
- Point of sale data
- Pharmaceutical prescription data

As a result of this wide range of data sources, it is possible to track a user's location via mapping apps, browser and search history, whom and what they "like" on social networks, videos and music they have streamed, their retail purchase history, the contents of their blog posts and online reviews, and much, much more.

Thus, the relation between artificial intelligence and big data is "bi-directional." Big data relies on artificial intelligence and machine learning to extract value from big datasets, and machine learning depends on the vast volume of data in order to learn.[32]

Big data, machine learning and artificial intelligence (AI) are enabling profitable commercial opportunities and social benefits through ***profiling*** and ***automated decisions***.

Profiling is the automated processing of personal data to evaluate, analyse or predict likely aspects of a person's interests, personal preferences, behaviour, performance at work, economic situation, health, reliability, location or movements.[33] Data analytics enables the identification of links between individuals and the construction of group profiles.[34]

Such inferences and predictions may be used for targeted advertising, or to make ***automated decisions*** (or to provide inputs to human decisions). Automated decisions are decisions made by computer processing systems without any human involvement (beyond the coding), typically based on inferences produced by profiling using machine learning models applied to big data. Inferences and predictions improve firms' ability to discriminate among consumers, offering them products and services suited to their preferences or needs, and at prices they are willing to pay. Examples include decisions whether to extend credit to an individual or to offer the person a job.

Numerous applications of big data and machine learning are being introduced in financial services, including:

- risk assessment, whether for lending or insurance, as discussed above, by companies such as Compare.com;[35]

[32] Artificial Intelligence, Robotics, Privacy and Data Protection, 38th International Conference of Data Protection and Privacy Commissioners, 2016, https://edps.europa.eu/sites/edp/files/publication/16-10-19_marrakesh_ai_paper_en.pdf.

[33] GDPR, Article 4(4) defines "profiling" as "any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements."

[34]

[35] https://www.compare.com/.

- investment portfolio management "robo-advisers" such as Betterment[36] and Wealthfront[37] that rely on algorithms to calibrate a financial portfolio to a consumer's investment goals and tolerance for risk;
- high-frequency trading (HFT) by hedge funds and other financial institutions such as Walnut Algorithms[38] and Renaissance Technologies[39] that use machine learning for making trading decisions in real time;[40]
- asset management, liquidity and foreign currency risk and stress testing;
- fraud detection by companies like APEX Analytics[41] and Kount[42] through detection and flagging of unique activities or behaviour anomalies to block transactions and for security teams to investigate; and
- a host of services such as security and digital identification, news analysis, customer sales and recommendations, and customer service.[43]

In some cases, these new uses are supported by legislation expressly authorising the use of artificial intelligence. For instance, Mexico's fintech reforms in 2018 amended the Securities Market Law to allow for special rules for automated advisory and investment management services (also known as robo-advisers).[44]

4.2 What is consumer protection?

Consumer protection is designed to protect humans where they are vulnerable. These may include protection of children, the elderly, and others who cannot protect themselves for physical or psychological reasons. It is widely acknowledged, though, that all consumers are vulnerable in some respects. We cannot know everything at all times. We have a limited ability to assess risk and benefits, i.e., we are subject to "bounded rationality."[45]

In consumer protection, the State intervenes through laws and processes in what would otherwise be a private relationship between consumer and provider. The need for this arises from perceived asymmetries between providers and consumers. These may include information asymmetries, where

[36] https://www.betterment.com/.
[37] https://www.wealthfront.com/.
[38] https://walnut.ai/en/.
[39] https://www.rentec.com/Home.action?index=true.
[40] https://www.quora.com/Why-are-machine-learning-neural-networks-and-other-AI-approaches-for-instance-not-more-widely-used-in-stock-market-predictions.
[41] https://www.apexanalytix.com/.
[42] https://www.kount.com/.
[43] Daniel Faggella, Machine Learning in Finance – Present and Future Applications, 18 September 2018, https://www.techemergence.com/machine-learning-in-finance/.
[44] Article 227 bis 1 of the Securities Market Law, Investment Advisors Chapter.
[45] See for example Cass Sunstein, Christine Jolls and Richard Thaler, *A Behavioural Approach to Law and Economics* Stanford Law Review 50 (1998); and Cass Sunstein and Richard Thaler, Nudge (2008) Yale University Press.

providers have greater data, knowledge and understanding than consumers. Differences in economic scale can also result in severe asymmetries of bargaining power. In addition, the transaction costs that consumers would face if they had to negotiate assurances about every product or service they acquire are too high to be feasible. As a result, a purely private, negotiated bargain between consumer and provider would be one-sided.

Consumer protection is formulated in various ways, but commonly seeks to promote the values of fairness, accountability and transparency (FAT).[46] The policy debate around consumer protection in relation to artificial intelligence and machine learning concerns the capacity of algorithms and machine learning systems to reflect such values.[47] Consumers may be vulnerable when dealing with services relying on computer processing for numerous reasons. Their functioning exceeds the comprehension of most of the population. Their precise, digital processes and results have a "seductive precision of output."[48] As a result, computers and results driven by them may be perceived as being objective and even fair. Today, however, there are risks that consumers will find some aspects of digital services to be unfair, unaccountable and non-transparent (the opposite of FAT), undermining trust between consumers and service providers and so hampering the prospects for growth in digital services.

Consumer protection laws typically involve the application of rules, principles and procedures to give consumers certain rights relating to the products and services they purchase. These rights include:

- rights prior to purchase ***(pre-engagement)***, such as information about the product or service provided;
- the provision, quality and functioning of the product or service itself ***(engagement)***; and
- post-purchase means of holding providers accountable (***post-engagement***).

The FAT values may apply in the pre-engagement phase, requiring notification to consumers about the product or service they are getting and sometimes securing express consent to it so that the consumer can take responsibility for their decisions.

However, a substantial part of consumer protection law operates on the premise that even if the consumer is notified about and consents to a product or service on the offered terms and conditions, such consent alone may not adequately achieve fairness, accountability and transparency. Thus, the FAT values may also apply in the engagement phase, i.e., to the actual product or service itself – its safety, quality or other features and conditions of provision. Therefore, consumer protection laws go further than pre-engagement notice and consent where notice and consent would not sufficiently protect the consumer and should not alleviate responsibility of the provider.

Again, FAT principles apply also in the post-engagement phase to ensure accountability mechanisms for securing explanations of why a given product or service was provided in the manner it was. They provide for consumers to have an opportunity to contest such decisions, and a means of redress where harm has resulted. Such protections may be applied regardless of whether the consumer has consented otherwise. For instance, many countries' laws do not permit consumers to submit to certain types of arbitration proceedings to resolve complaints and to bargain away their rights to be heard in court.

---

[46] For instance, section 5 of the US Federal Trade Commission Act prohibits "unfair or deceptive acts or practices in or affecting commerce," which has been one of the foundations of digital privacy enforcement in the USA.

[47] Kate Crawford *et al*., The AI Now Report: The Social and Economic Implications of Artificial Intelligence Technologies in the Near Term 6-8 (2016), https://artificialintelligencenow.com/media/documents/AINowSummaryReport_3_RpmwKHu.pdf.

[48] Paul Schwartz, *Data Processing and Government Administration: The Failure of the American Legal Response to the Computer*, 43 HASTINGS L.J. 1321, 1342 (1992).

Instead, such laws insist on procedures ensuring that consumers have a fair and transparent process to hold providers accountable.

Thus, many countries' laws protect consumers against misleading product descriptions, unfair contract terms (e.g., exclusion of liability), faulty products and lack of redress mechanisms. Such laws prohibit manufacturers and retailers from negotiating such terms with consumers, so that they cannot argue that consumers consented to them when they bought the product or service. The consumer protection approach introduces minimum common standards and procedures to provide a base level of protection rather than leaving everything to consumer autonomy and responsibility.

Consumer protection laws have an important, even symbiotic, relationship with competition law and policy. The asymmetry of bargaining power that justifies consumer protection may be exacerbated where a market is concentrated and consumers lack alternatives for a given service. There are currently increasingly calls to address high levels of market concentration in data markets from a competition policy perspective. The European Commission and several Member States have been developing theories of harm around large tech firms that gather consumer data through business models that use such data to generate advertising revenue. Some authorities such as Germany's competition authority, the Bundeskartellamt, have raised the possibility that failure to respect consumer privacy rights can in some circumstances amount to abuse of dominant market position under competition law. The focus of this paper, however, is not on competition law aspects of big data and machine learning, but on consumer protection and privacy issues.

A number of consumer protection measures discussed in this paper are just as pertinent to sole proprietor businesses and mico-, small- and medium-sized enterprises (MSMEs). Where countries' laws do not treat these as data subjects or consumers, they may not benefit from the protections afforded under data protection and privacy laws. There are strong arguments in favour of extending such protections to such businesses.

4.3 What is data privacy?

Privacy risks

Not all big data and machine learning techniques rely on personal data or give rise to consumer protection issues. There is extensive data that does not relate to an identifiable person that can be used for commercial and social benefits. However, where personal data is used, it may give rise to concerns about the privacy of the individuals concerned.

Privacy encompasses a broad range of notions. Whether viewed as a value or in terms of rights or protections, it has been boiled down by some scholars to concerns about "individuality, autonomy, integrity and dignity,"[49] part of a broader range of ideas concerning freedom in personal and family life.

While privacy may refer to the individual's freedom from others interfering with personal choices, particularly relating to their body, a large part of privacy concerns what is known by whom about the individual, and thus treatment of personal data. Data privacy is not the same as data security. Secure management of data is necessary to protect privacy, but privacy concerns specific values relating to individual persons that need to be taken into account when ensuring data is secured.

[49] Lee A Bygrave, Data Protection Law: Approaching Its Rationale, Logic and Limits (Kluwer Law Intl 2002) 128–129.

Thus in the digital context, privacy involves controls on the collection, use and sharing of personal data. "Personal data" is a term with a potentially vast meaning, extending to any information relating to an identifiable individual.[50] Most data protection regimes recognise that some personal data is more sensitive or easily susceptible to abuse than others and apply tightened controls accordingly.

Data about a person may be:

- provided by the person (e.g., a user name, or a postcode);
- observed about the person (e.g., location data); or
- derived from provided or observed data (e.g., country of residence derived from the postcode); or
- inferred from the foregoing (e.g., a credit score) through deduction or reasoning from such data.[51]

Consumers face privacy risks where their personal data may be accessed by unauthorised users, may be abused, or may be used for profiling that leads to subjective inferences about the consumer that may be difficult to verify, and may result in automated decisions that affect the individual's life.

A key privacy risk relates to the aggregation of personal data. In the case of big data, this risk is aggravated where personal data is not anonymised, or where pseudonymisation or anonymisation has been attempted but the re-identification of the person remains possible (see section 6.3). Increasingly, countries are legislating to protect the personal data and privacy of their subjects, with an important theme being the minimisation of data collection, use and sharing.

The scope of the personal data that may be generated and shared may, as a result of big data and machine learning, include inferences made about them and predictions of their behaviour. However, inferences about a person made from their personal data are typically not treated as personal data to be protected.[52] Laws often restrict privacy protections to rectifying, blocking or erasing the personal data that is input into algorithms, but not to the evaluation of that data or decisions based on such evaluation. As recently suggested in relation to the GDPR, "Ironically, inferences receive the least protection of all the types of data addressed in data protection law, and yet now pose perhaps the greatest risks in terms of privacy and discrimination."[53]

Protecting privacy

Potential data protection remedies include the consumer's right to know what personal data is collected,[54] the right to rectify inaccurate personal data and to complete incomplete personal data,[55] the right to have personal data deleted,[56] the right to port data to a third party,[57] and the right to object to processing of personal data (including for profiling).[58] While the European Union has adopted all of

[50] The GDPR defines "personal data" in Article 4(1) as "any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person."

[51] The EU's Article 29 Working Party distinguished between these three categories in Article 29 Data Protection Working Party, 'Guidelines on Automated Individual Decision Making and Profiling for the Purposes of Regulation 2016/679' (n19) 8, available at http://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053.

[52] Sandra Wachter & Brent Mittelstadt, *A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI*, (2019) Colum. Bus. L. Rev. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3248829 (Wachter & Mittelstadt).

[53] *Ibid.*

[54] GDPR, Articles 13-15.

[55] GDPR, Article 16.

[56] GDPR, Article 17.

[57] GDPR, Article 20.

[58] GDPR, Article 21.

these remedies in the GDPR, many countries focus more on rights of access and rectification and breach notification obligations.

Data protection and privacy are not the domain solely of high income, northern hemisphere countries. Today, 107 countries, of which 66 are developing or transition economies, have adopted laws on data protection and privacy, and more are on the way.[59] Many countries outside Europe have committed to stringent levels of data protection by signing Convention 108 (for instance, Mexico signed in 2018).

EU's GDPR not only provides reinforced rights and obligations, but has significant extraterritorial impact. The GDPR requires that personal data be protected when it is exported to and processed in countries outside Europe. It applies to the processing of any individual's data who is "in the Union" even if the data processing occurs outside the EU. Thus, countries dealing with Europe in digital services and non-European companies who are likely to process data of Europeans must adopt GDPR-like protections. For instance, Japan completed discussions to establish data protection and privacy regimes sufficiently similar to the EU to merit "adequacy" treatment in 2018, and talks are ongoing with South Korea. Uruguay was previously granted adequacy in 2012 under the EU's prior data protection directive regime.

Some countries treat data protection and privacy as a matter of constitutional law. Mexico's Constitution, for example, prohibits intrusion onto an individual's person, family, domicile, documents or belongings (including any wiretapping of communication devices), except when ordered by a competent authority supported by the applicable law.[60] The right to data protection is provided for, setting a standard for all collecting, using, storing, divulging or transferring (collectively processing) of personal data to secure the right to privacy and self-determination.[61]

India's Supreme Court in 2017 declared privacy a "fundamental right," protected by the Constitution,[62] echoing the United States[63], the European Union[64] and numerous other jurisdictions. In some cases, these matters have a specific written foundation in the Constitution itself. Brazil's Constitution, for example, has a right of "*habeas data*" that gives individuals the right to access and correct personal data about themselves held by public agencies.[65] Some countries, such as Kenya, have a constitutional right of privacy but have not (as yet) introduced stand-alone legislation.

---

[59] UNCTAD, Global cyberlaw tracker, as of 217 September 2018. Another measure put the number at 120 in 2017. See Greenleaf, Graham, Global Data Privacy Laws 2017: 120 National Data Privacy Laws, Including Indonesia and Turkey (January 30, 2017). (2017) 145 Privacy Laws & Business International Report, 10-13; UNSW Law Research Paper No. 17-45. In Africa alone, 22 countries already have privacy and data protection laws: Angola (2016), Benin (2009), Botswana (2018), Burkina Faso (2004), Chad (2015), Cape Verde (2001), Côte d'Ivoire (2013), Equatorial Guinea (2016), Gabon (2011), Ghana (2012), Lesotho (2012), Madagascar (2014), Mali (2013), Mauritius (2017), Mauritania (2017), Morocco (2009), Senegal (2008), Seychelles (2002), South Africa (2013), Tunisia (2004), Zambia, and Zimbabwe (2003). Algeria, Democratic Republic of the Congo, Ethiopia, Kenya, Malawi, Mauritania, Niger, Nigeria, Rwanda, Sierra Leone, Swaziland, Tanzania and Uganda, have prepared and are considering draft legislation. See CIPESA, State of Internet Freedom in Africa 2018: Privacy and Data Protection in the Digital Era - Challenges and Trends in Africa, September 2018 at 7.

[60] Paragraphs 1 and 12 of Article 16 of the Mexican Constitution.

[61] Paragraph 2 of Article 16 of the Mexican Constitution.

[62] *K.S. Puttaswamy & Anr. v. Union of India & Ors*. (2017) 10 SCC 1.

[63] In 1974, the US Congress stated in the federal Privacy Act that "the right to privacy is a personal and fundamental right protected by the Constitution of the United States."

[64] In December 2009, when the Lisbon Treaty took force, the EU's Charter of Fundamental Rights guaranteed privacy and data protection as among 50 other fundamental rights.

[65] Article 5 (LXXII), Constitution of the Federated Republic of Brazil, 3rd edition, 2010, http://english.tse.jus.br/arquivos/federal-constitution.

The proliferation of data and the potential for big data technologies to violate privacy recently led the Indian Supreme Court to limit the use of Aadhaar, India's national digital ID system.[66] The Court ruled that requiring use of Aadhaar for services other than public services like social payments, including mandatory use of Aadhaar for know-your-customer (KYC) in banking and telecommunications, would be unlawful.[67] The Court found that specific legal requirements to link the Aadhaar system with all new and existing bank accounts and mobile phone numbers violated the fundamental right to privacy. It would enable "commercial exploitation of an individual['s] biometric and demographic information by private entities."

Treating privacy as a fundamental right is only one approach to ensuring the protection of users. Some countries regard privacy less as a matter of fundamental rights and more as a matter of consumer protection. While this may result in a weaker commitment to general privacy protection, it may result in greater focus on the trade-offs and cost-benefit issues involved in regulating to protect privacy. Consumer protection agencies will more often have to carry out a balancing act when considering whether a given conduct is unfair to consumers and should be viewed as unlawful.[68]

This approach does not prevent focused privacy law and regulation where it is most important, which in most countries has included the health, financial and communications sectors, and protection of children. Some countries have no generally applicable privacy law, but have developed substantial privacy law and regulation separately in such individual sectors at different times and without strong coordination among the sectoral legal provisions. While this may allow privacy concerns to be tailored to a given sector's specificities, it also risks creating complexity, inconsistencies among sectors and challenges to harmonisation across borders.

Some countries have preferred to establish non-binding standards for privacy protection, such as China's National Standards on Information Security Technology – Personal Information Security Specification GB/T 35273-2017 entered into effect in 2018. This establishes numerous standards for protecting personal information, loosely based on Europe's GDPR. It sets out practices that regulators will expect to see introduced when they audit firms and enforce China's existing data protection laws, in particular the 2016 Cybersecurity Law. Further national standards including on big data and data anonymisation, are expected to be introduced.

Even jurisdictions that assert privacy as a fundamental right recognise the necessity of weighing the individual's interest against the interest of public and private organisations, and broader social interests such as scientific research, innovation, national security and crime enforcement. Not only is there in many jurisdictions a basic right to conduct a business,[69] there may be intellectual property and trade secrets rights involved as well.

Protecting privacy, like any regulation, involves costs, such as the financial costs of compliance and the opportunity costs of new services relying on access to personal data. Some argue that such costs are a

[66] See https://uidai.gov.in/ for more information about Aadhaar.
[67] K.S. Puttaswamy & Anr. v. Union of India & Ors. (2018), Paras 159-160, https://www.sci.gov.in/supremecourt/2012/35071/35071_2012_Judgement_26-Sep-2018.pdf.
[68] The US Federal Trade Commission, the general privacy regulator, is subject to a statutory balancing act as follows: "The Commission shall have no authority under this section … to declare unlawful an act or practice on the grounds that such act or practice is unfair unless the act or practice causes or is likely to cause substantial injury to consumers which is not reasonably avoidable by consumers themselves and not outweighed by countervailing benefits to consumers or to competition." 15 U.S. Code 45(n).
[69] E.g., Article 16 of the EU Charter of Fundamental Rights.

justifiable economic investment because strengthened trust will increase demand for services. Some view such investments, as Tim Cook, CEO of Apple recently put it, as a choice of what kind of society we want to live in.[70]

In any scenario, it is reasonable and appropriate for legislators and regulators to consider not only the ideal of privacy but the impediments to innovation and productive purposes, and the diversion of resources, that compliance-focused protections may entail. It is prudent to identify and quantify as best possible the benefits and the costs, and prioritise risks that are most harmful. As the World Bank and Consultative Group to Assist the Poor[71] (CGAP) put it, "[p]olicy makers face the challenge of striking the right balance between promoting the benefits of the expanded use of alternative data while ensuring adequate data protection and attention to consumer privacy across the eco-system."[72]

5 Pre-engagement: notice and consent

This section considers the requirement in many data protection and privacy laws to notify the consumer of the fact that, and purpose for which, their personal data will be collected, used and shared with third parties, and to obtain their consent – before they engage in submitting data and requesting the service.

5.1 Notice and consent requirements

An increasing number of countries' data protection laws and standards provide for stringent regulation of collection, use and sharing of data. These require firms to inform consumers when they are collecting personal data about them, and of the purpose for which the data will be processed, as well as whether they may transfer the data to third parties.[73] Third parties may also be required to notify a consumer where they acquire personal information about the consumer.[74] This is rarely required, however, and even when it is, it may be restricted to categories of information and not inferences about the individual.

Two longstanding themes of data protection and privacy law are "purpose specification" and relatedly "data minimisation": the requirement to specify the purpose for which data is collected, used and shared, and to limit collection, use and sharing to data which is relevant, adequate and necessary for (or proportionate to) that purpose.[75] As any collection and use of data may increase risk to security and privacy, the objective is to minimise or avoid additional risk beyond what is necessary for the purpose. This aims to prevent "function creep" whereby data that is originally collected for one purpose is then

[70] Speech given to the International Conference of Data Protection and Privacy Commissioners (ICDPPC) in Brussels on 24 October 2018. Complete transcript available at https://www.computerworld.com/article/3315623/security/complete-transcript-video-of-apple-ceo-tim-cooks-eu-privacy-speech.html.
[71] CGAP is an arm of the World Bank focussed on alleviating poverty through financial inclusion. See https://www.cgap.org/about/governance.
[72] World Bank & CGAP, Data Protection and Privacy for Alternative Data, GPFI- FCPL SUB-GROUP DISCUSSION PAPER -DRAFT-MAY,4 2018 p5.
[73] GDPR, Article 13. China's Personal Information Security Specification 2018 requires data subjects to be informed about the scope, purpose and rules of the processing of their personal information in an explicit, comprehensible and reasonable manner.
[74] GDPR, Article 14.
[75] For example, China's Personal Information Security Specification of 2018 provides that, unless the data subject otherwise agrees, a personal data controller should limit the processing of personal information to what is necessary to accomplish a specified purpose and delete such information as soon as the purpose is fulfilled.

used for other purposes.[76] The OECD Use Limitation Principle, for instance, refers to the need to obtain consent from the individual if the data is to be used for purposes other than the original purpose for which it was collected.[77]

There are sometimes exceptions to notice and consent rules that allow for uses of data beyond its initial purpose of collection, such as for statistical purposes or when it will be used for scientific research.[78] These often depend on large datasets for the same reason that machine learning does generally. There are potential grey areas between what comprises statistical purposes or scientific research and what constitutes product development in the provision of financial services. However, these exceptions to the purpose specification and data minimisation rules are typically not wide in scope.

Many countries' laws, and international and regional standards also require the individual to "opt in" by providing consent to collection, use and sharing of personal data.[79] Where this is not required or obtained, some jurisdictions allow the individual to "opt out" by providing notice that they do not wish their personal data to be collected, used or shared with third parties.[80] When the consumer is not provided with a choice, data protection laws may impose obligations of transparency, requiring data controllers to provide clear and accessible explanations in privacy policies as to how and for what purpose their data will be used and shared.[81]

When it comes to decisions made as a result of big data and machine learning, one approach is simply to outlaw them where they pose unacceptable risk. This has been recommended, for instance, for the use of lethal weapons. With very limited exceptions, automated cars are not yet allowed on the streets, although laws are being developed to enable these.

However, recognising that many automated processes can bring benefits to consumers, these are often permitted so long as consumers are notified of the automated decision-making and have an opportunity to opt out. For instance, the GDPR requires notice of "the existence of automated decision-making, including profiling, […] and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject."[82] It also provides in Article 22(1) that individuals "shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her."[83]

---

[76] See generally, for example, Els J. Kindt, *Privacy and Data Protection: Issues of Biometric Application, A Comparative Analysis*, Heidelberg, Dordrecht, New York, London: Springer, 2013.

[77] OECD, *Guidelines on Protection of Privacy and Cross-Border Flows of Personal Data*, as amended in 2013, Principle 10, http://oecdprivacy.org/.

[78] GDPR, preamble paragraph 50 and Article 5(1)(b).

[79] For example, the *2016 EU General Data Protection Regulation* states in its Preamble at para 40: "In order for processing to be lawful, personal data should be processed on the basis of the consent of the data subject concerned or some other legitimate basis, laid down by law, either in this Regulation or in other Union or Member State law. . ." (emphasis added). Consent means "any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her." GDPR, Article 4(11).

[80] E.g., California Consumer Privacy Act 2018.

[81] E.g., the *2004 APEC Privacy Framework* requires data controllers to provide "clear and easily accessible statements about their practices and policies with respect to personal information."

[82] GDPR, Articles 13, 14 and 15.

[83] Likewise, Kenya's Data Protection Bill being considered for enactment provides in Section 31, "Every data subject has a right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning or significantly affects the data subject." It also has exceptions to this, including whether the automated processing is necessary for a contract, authorized by law with safeguards and based on explicit consent.

This opt-out right may be helpful, but it only goes so far. Automated decisions are permitted under the GDPR where necessary to enter into a contract with the individual, or with their consent.[84] Where new services rely on profiling to establish eligibility, and are expected to be made rapidly, often remotely and electronically, automated decisions may be necessary to enter into the contract. And where an individual's need or desire for a product or service exceeds their personal intolerance for being the subject of automated processing, a binary choice is presented and the individual may have no meaningful option but to consent.

5.2 The context of big data

Big data and machine learning pose challenges to the notice and consent approach to data protection and privacy law and regulation.

Purpose specification in the context of machine learning

Complying with notice requirements involves providing to individuals a detailed specification of the purpose of collecting their personal data, and closely monitoring operations to avoid exceeding such purpose. Machine learning detects patterns and then delves into deeper layers, identifying further patterns, and these may reveal use cases which may not directly relate to the original purpose of data mining. As a result, the purpose for which the data may end up being used may not be known at the time the data is being collected, or when consent is obtained. Only vague purposes may be identifiable at that time, which indeed accounts for the generally vague nature of privacy policies and data collection notifications.

Data minimisation in the context of big data

In addition, as machine learning techniques are more effective in detecting patterns in larger datasets over time, the very nature of big data is to collect the maximum possible amount of data – and to retain it for as long as possible.

Thus, the very notion of data minimisation (to collect as little data as possible and hold it for as short a time as possible according to the purpose for which it was collected) runs counter to the modus operandi of the industry. It undermines the prospects for genuinely informative notification to users of the purpose of collection. Disclosures, monitoring and compliance may also be difficult and expensive. Describing the purpose as very broad in order to avoid such limits may well not be legally acceptable. A 2014 report to the US President suggested that "The notice and consent is defeated by exactly the positive benefits that big data enables: new, non-obvious, unexpectedly powerful uses of data."[85]

Limits of consumer responsibility

Furthermore, despite efforts to make notifications simple and understandable, such documents are not frequently read and understood by the consumer.[86] This undermines the notice and consent approach,

[84] GDPR, Article 22(2).
[85] President's Council of Advisors on Science & Technology, *Big Data and Privacy: A Technological Perspective*, The White House, May 1, 2014.
[86] See, e.g., Whitley, E. A., and Pujadas, R. (2018). Report on a study of how consumers currently consent to share their financial data with a third party, *Financial Services Consumer Panel* at ii. "The evidence from the empirical research suggests that consent is frequently neither freely given, nor unambiguous nor fully informed. Over half of the contributors claimed not to read any terms and conditions for products and services that they sign up for, including the specific services that access their financial data. Similarly, only a small proportion of participants correctly answered a question about a detail in the policy even after having an opportunity to re–read the policy in a research setting."

further rendering it not only ineffective but misleading, often displacing onto the consumer a burden that they are unable to bear, and creating a perception of legitimacy which is not justified. Privacy policies and consent may "check the box" as part of a compliance-oriented approach, but they do little substantively to enable consumers to understand how their data may be used and shared with third parties, let alone the implications of such use and sharing.[87]

Some have suggested simplifying notices because artificial intelligence and machine learning design specifications are currently incapable of providing satisfactory accountability and verifiability, making them more impactful – like "skull and crossbones found on household cleaning supplies that contain poisonous compounds."[88]

In addition to the difficulty of expecting the consumer to bear the burden of responsibility for matters that are often beyond their comprehension, the manner by which consent is solicited on a binary take-it-or-leave-it basis accentuates the problem.

Privacy in context

Some have suggested that one approach is to recognise that privacy is generally very context-specific, relating to the expectations that a person would reasonably have in light of the nature of the situation or transaction. An individual might expect high levels of privacy (confidential treatment) when dealing with medical, financial or other personal matters, but be quite relaxed about being overheard in a public square, or being offered assistance in searching for products in a shop. One might have different expectations regarding privacy when carrying out research depending on the context, including the subject matter or purpose of the research. Similarly, whether one might expect to be able to enjoy entertainment in private may depend on the nature of the content.

It has been suggested, therefore, that "contexts, not political economy, should determine constraints on the flow of information," so that privacy protections online should be aligned with such expectations.[89] This might mean tighter restrictions on collection, use and sharing of personal data in some situations even if notice and consent are provided. The Consumer Privacy Bill of Rights proposed by President Obama's White House in 2012[90] sought to take this approach, adopting as its third principle, "Respect for Context," which was explained as the expectation that "companies will collect, use, and disclose personal data in ways that are consistent with the context in which consumers provide the data."[91]

To the extent that user consent continues to be viewed as a legitimate basis for collecting and using data, improvements may be made to the means by which consent is obtained. In addition to improving the plain language of notifications, such improvements may include using tiered consent which differentiates between types of data according to the types of purpose for which it may be used or which types of organisation may use it. Sunset clauses for consent to expire may also be appropriate.[92]

---

[87] Daniel J. Solove, *Privacy Self-Management and the Consent Dilemma*, 126 HARV. L. REV. 1880, 1889-93 (2013).
[88] IEEE Global Initiative (see footnote 217) at p159.
[89] Helen Nissenbaum, *Privacy in Context: Technology, Policy, and the Integrity of Social Life,* Stanford University Press (2010); and Helen Nissenbaum, A Contextual Approach to Privacy Online, Daedalus (2011). https://www.amacad.org/publications/daedalus/11_fall_nissenbaum.pdf.
[90] https://obamawhitehouse.archives.gov/sites/default/files/privacy-final.pdf.
[91] Consumer Data Privacy in a Networked World: A Framework for Protecting Privacy and Promoting Innovation in the Global Digital Economy (2012) http://btlj.org/2012/03/president-obamas-privacy-bill-of-rights-encouraging-a-collaborative-process-for-digital-privacy-reform/.
[92] See, e.g., Bart Custers, *Click Here to Consent Forever: Expiry Dates for Informed Consent*, Big Data & Society, January–June 2016: 1–6. http://journals.sagepub.com/doi/10.1177/2053951715624935.

Technologies of consent management

Efforts are also being made to develop technologies and services to manage consent better. This relies on using forms of digital rights management, attaching permissions to personal data, and enabling automated negotiations between individuals and those who receive their data concerning its collection, use and sharing. Such approaches seek to improve transparency and consumer control, and thus also to make data more freely available due to increased trust.[93] Instead of binary consent decisions whereby consumers either grant access to all of their data or they cannot enjoy the service, there may be ways to allow graduated consent according to preferences for sharing and storing personal data.

Making this possible on a large scale may require use of algorithmic tools acting as an agent,[94] guardian or fiduciary – "algorithmic angels"[95] – on behalf of the consumer. Some have suggested that providers of such personal data management services could inform and educate individual consumers and "negotiate" on their behalf, suggesting how requested data could be combined with other previously provided data, inform the consumer if data is being used in a manner that was not authorised, or make recommendations to the consumer based on their profile.[96] Such a process could even involve setting terms for the sharing of data, including payment to the consumer, or retraction of previously granted consent if the conditions of such consent were breached.

There appears to be a genuine commercial opportunity for investment and innovation to improve management of such consumer consent. Firms like Sudo[97] allow consumers to make easy use of a pseudonym for a variety of digital interactions, from telephone calls to e-commerce and online dating.

Related ideas involve the consumer generally having greater control over their data. For instance, India's "Digital Locker," which is part of the India Stack, enables individuals to have greater control over who may access their data, including creating an auditable record of when their records are accessed. Other ideas include conceiving of a property right of ownership over personal data, although this has approach not yet gathered steam.

All of these suggestions aim to enhance consumer control over personal data, reducing the currently prevailing asymmetries. There may even be benefits to the quality of data that is gathered as a result. Some have suggested that allowing individuals to set their preferred level of anonymity when responding to requests for data gathering (e.g., for post-purchase consumer feedback or in health surveys) may improve the reliability of data submitted.[98]

---

[93] Pentland Alex (MIT). Big Data's Biggest Obstacles. 2012. Available at: https://hbr.org/2012/10/big-datas-biggest-obstacles.

[94] Work is underway on a standard for an "AI agent" under IEEE project *P7006 - Standard for Personal Data Artificial Intelligence (AI) Agent*, https://standards.ieee.org/project/7006.html.

[95] Jarno M. Koponen, *We need algorithmic angels*, TechCrunch 2014, https://techcrunch.com/2015/04/18/we-need-algorithmic-angels/.

[96] See Ethically Aligned Design, at footnote 217 at p103. See also Mike Orcutt, *Personal AI Privacy Watchdog Could Help You Regain Control of Your Data*, MIT Technology Review, 11 May 2017, https://www.technologyreview.com/s/607830/personal-ai-privacy-watchdog-could-help-you-regain-control-of-your-data/, and the related Privacy Assistant mobile app, https://play.google.com/store/apps/details?id=edu.cmu.mcom.ppa&hl=en.

[97] https://mysudo.com/.

[98] Kok-Seng Wong & Myung Ho Kim, *Towards a respondent-preferred k_i-anonymity model*, Frontiers Inf Technol Electronic Eng (2015) 16: 720. https://doi.org/10.1631/FITEE.1400395. "The level of anonymity (i.e., *k*-anonymity) guaranteed by an agency cannot be verified by respondents since they generally do not have access to all of the data that is released. Therefore, we introduce the notion of k_i-anonymity, where *k* is the level of anonymity preferred by each respondent *i*. Instead of placing full trust in an agency, our solution increases respondent confidence by allowing each to decide the preferred level of protection. As such, our protocol ensures that respondents achieve their preferred k_i-anonymity during data collection and guarantees that the collected records are genuine and useful for data analysis."

6 Engagement: operations

This section discusses engagement: the consumer's experience with big data and machine learning, and conversely the collection, use, storage and transfer of the consumer's data by big data and machine learning firms. Sections 6.1 and 6.2 consider consumer concerns and legal issues that arise from the substantive results of the data processing, in particular responsibility for accuracy and biased decision-making. Section 6.3 considers protections for consumers against the risk of the release of their data through data breach and re-identification, focusing on the techniques of de-identification, pseudonymisation and anonymisation. Section 6.4 turns to the risks to consumers that arise through transfers of data in the vibrant data broker market, and increased regulation of this market segment.

6.1 Accuracy

Accuracy of data inputs

The successful functioning of machine learning models and accuracy of their outputs depends on the accuracy of the input data. Some of the vast volumes of data used to train the system may be "structured" (organised and readily searchable) and some may be "unstructured."[99] The data may have been obtained in different ways over time from a variety of sources, some more and some less directly. The wider the net of data that is collected, the greater the chances are that data will be out of date and that systematic updating processes are not applied. Historical data may have even been incorrect from the start.

These factors may result in questionable accuracy of data inputs to the algorithms. This may be true both for the personal data about the individual who is the subject of an automated decision (to which the machine learning model is applied), as well as for the wider pool of data used to train the machine. If the training data is inaccurate, the model will not function to produce the intended outputs when applied to the individual's personal data. All of these problems may give rise to erroneous inferences about the consumer.

Data protection and privacy laws thus increasingly set some form of legal responsibility on firms to ensure the accuracy of the data they hold and process. Mexico's data protection legislation applies a quality principle requiring data controllers to verify that personal data in their databases is correct and updated for the purposes for which it was gathered. [100]

This raises the question about the accuracy of data in the wider data ecosystem, and the extent to which firms should be held responsible for inaccuracy or to contribute to accurate information more broadly.

Responsibility for data accuracy in financial services

Sector-specific laws governing financial services often emphasise the importance of ensuring accuracy of data used for financial services. Data used for credit scoring is an example.[101] Credit reporting

[99] Structured data has a high degree of organization, such that inclusion in a relational database is seamless and readily searchable by simple, straightforward search-engine algorithms or other search operations (e.g., payment and transaction reports). Unstructured data either does not have a pre-defined data model or is not organized in a predefined manner (e.g., social media entries, emails and images).
[100] Similarly, GDPR Article 5(1)(d) provides, "Personal data shall be accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay."
[101] The OECD Data Quality Principles (Principle 8), the APEC Privacy Framework (Principle 21), the Madrid Resolution Data Quality Principle, and Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (referred to as Convention 108) (Article 5) all include principles requiring that information be accurate and up-to-date. The G20 High-Level Principles

bureaus are typically subject to regulation and strong internal controls to ensure accuracy of the data they hold on individuals. Such credit reporting systems reduce the costs of lending by reducing risk (and thus loan default losses, provisioning for bad debt, and need for collateral) inherent in information asymmetries between lenders and borrowers. They provide lenders with information to evaluate borrowers, allowing greater access to financial services.[102] Because of the importance of their data in credit and other decision-making, credit reference bureaus provide individuals with a means of correcting inaccurate information.

However, this formal information system is now only part of a wider data-rich environment, most of which is not regulated. The advent of big data and machine learning poses a risk that existing legislation and policy guidance does not keep up with the data-rich environment. For instance, the first principle of the World Bank's General Principles on Credit Reporting (GPCR), published in 2011[103], is that "credit reporting systems should have relevant, accurate, timely and sufficient data – including positive – collected on a systematic basis from all reliable, appropriate and available sources, and should retain this information for a sufficient amount of time."

Questions arise about how exactly this sort of policy guidance should apply today – just eight years later – to information about individuals supplied and collected for the purpose of making credit decisions. Big data and machine learning may collect and use data that varies greatly in its relevance, accuracy and timeliness.

These challenges apply also to laws that were written before the advent of big data and machine learning and even the internet itself. Firms that do not consider themselves to be credit reference bureaus may nevertheless find themselves subject to legal obligations that apply to traditional credit reference bureaus. In some cases, such companies could find themselves subject to claims for failure to supply accurate information that has a bearing on a person's credit worthiness.

Many countries recognise a public interest in ensuring "fair and accurate credit reporting," as formulated in the US, for example. [104] This both benefits the functioning of financial services markets and protects consumers. For this reason, consumer reporting agencies whose data are used for credit transactions, insurance, licensing, consumer-initiated business transactions, and employment are often regulated.[105]

However, many countries' consumer reporting laws were enacted before the advent of the internet, let alone big data and machine learning. Some countries have a broader concept of consumer reporting agencies. In the US, for example, the Fair Credit Reporting Act (FCRA) applies to companies that regularly disseminate information bearing on an individual's "credit worthiness, credit standing, credit

for Digital Financial Inclusion (HLP-DFI) calls for the development of "guidance to ensure the accuracy and security of all data related to: accounts and transactions; digital financial services marketing; and the development of credit scores for financially excluded and underserved consumers. This guidance should cover both traditional and innovative forms of data (such as data on utility payments, mobile airtime purchases, use of digital wallet or e-money accounts, social media and e-commerce transactions)." https://www.gpfi.org/sites/default/files/G20%20High%20Level%20Principles%20for%20Digital%20Financial%20Inclusion.pdf

[102] International Finance Corporation (IFC), Credit reporting knowledge guide. Washington, DC. http://www.ifc.org/wps/wcm/connect/industry_ext_content/ifc_external_corporate_site/industries/financial+markets/publications/toolkits/credit+reporting+knowledge+guide.

[103] http://www.worldbank.org/en/topic/financialsector/publication/general-principles-for-credit-reporting.

[104] §1681(a)(1).

[105] §§1681a(d)(1)(A)–(C); §1681b.

capacity, character, general reputation, personal characteristics, or mode of living."[106] The FCRA requires consumer reporting agencies to "follow reasonable procedures to assure maximum possible accuracy" of consumer reports; to notify providers and users of consumer information of their responsibilities under the Act; to limit the circumstances in which such agencies provide consumer reports "for employment purposes"; and to post toll-free numbers for consumers to request reports. It also creates liability for failure to comply with these requirements.[107]

In a 2016 report, the US consumer agency, the Fair Trade Commission (FTC), considered how big data is used in credit reporting decisions.[108] The FTC clarified that data brokers that compile "non-traditional information, including social media information" may be considered to be credit reporting agencies subject to these obligations.

This is not a mere theoretical possibility. For instance, in the recent US Supreme Court case *Spokeo v Robins*,[109] Spokeo operated a website which searched and collected data from a wide range of databases. It provided individuals' addresses, phone numbers, marital status, approximate ages, occupations, hobbies, finances, shopping habits and musical preferences and allowed users to search for information about other individuals. The plaintiff, Robins, alleged that Spokeo incorrectly described him as a wealthy, married professional, resulting in him being adversely perceived as overqualified for jobs. Robins claimed that Spokeo was a "consumer reporting agency" under the FCRA,[110] and was liable to him for having supplied incorrect information. The case was resolved on other grounds, but the potential breadth of such legacy legislation poses challenges for firms operating in the data business. It may give rise to responsibilities to consumers for accuracy of data used to make credit and other decisions that were not anticipated, weaken legal certainty and undermine business innovation and investment.

Credit reporting requirements and the wider information ecosystem

The discussion above concerned the responsibilities to consumers that firms may have when dealing with data in non-traditional ways, in particular regarding the accuracy of data they use for decisions in financial services. A related question arises concerning firms' responsibility to contribute to the wider information ecosystem that is traditionally regulated by disclosure and reporting obligations.

Disclosure obligations arise in numerous contexts, whether due to securities laws requirements applicable to public companies, health and safety disclosures for medicines, or consumer products that pose particular risks. In the financial services context, for example, a person's credit history is useful

---

[106] §1681a(d)(1).
[107] §1681e(b); §1681e(d); §1681b(b)(1); §1681j(a); and §1681n(a).
[108] Federal Trade Commission, *Big data: A tool for inclusion or exclusion? Understanding the issues*. Washington, DC (2016). https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf.
[109] 136 S. Ct. 1540 (2016).
[110] Fair Credit Reporting Act of 1970, 84 Stat. 1127, as amended, 15 U. S. C. §1681 et seq. The US Fair Credit Reporting Act (FCRA) seeks to ensure "fair and accurate credit reporting." §1681(a)(1). It regulates the creation and the use of "consumer report[s]" by "consumer reporting agenc[ies]" for credit transactions, insurance, licensing, consumer-initiated business transactions, and employment. §§1681a(d)(1)(A)–(C); §1681b. The FCRA was enacted long before the Internet, and applies to companies that regularly disseminate information bearing on an individual's "credit worthiness, credit standing, credit capacity, character, general reputation, personal characteristics, or mode of living." §1681a(d)(1) The FCRA requires consumer reporting agencies to "follow reasonable procedures to assure maximum possible accuracy of" consumer reports; to notify providers and users of consumer information of their responsibilities under the Act; to limit the circumstances in which such agencies provide consumer reports "for employment purposes"; and to post toll-free numbers for consumers to request reports. It also creates liability for failure to comply with these requirements. §1681e(b); §1681e(d); §1681b(b)(1); §1681j(a); and §1681n(a).

data for a financial service provider, reducing the asymmetry of information between lender and borrower. In order to improve competition among service providers that hold such data and the functioning of financial markets, some financial service providers are often required to report credit data about consumers to consumer reporting organisations which organise and make it available to the market as a whole.

In many countries, only banks (i.e., entities that are regulated, typically with banking licences, for deposit taking, lending and other related activities) are required to report to credit reference bureaus for inclusion in the credit reference bureau's records and analytics. Today, the question arises whether non-banking financial service providers that rely on automated decisions using alternative data to profile risk should be obligated to report the results of such lending to credit reference bureaus as well.

Some consider that alternative lenders should be required to supply credit data to credit reference bureaus about a consumer's loan that is successfully repaid (positive reporting data) as well as where the consumer defaults on the loan (negative reporting data).[111] Doing so may provide a more "level playing field" of regulatory obligations for similar activities (lending) rather than applying different regulatory obligations depending on the type of entity (a bank as opposed to a non-bank). This may also increase the broader range of data available about consumers, and so enrich and plug gaps in the data ecosystem.

These potential advantages need to be weighed in light of how the alternative credit market is developing. Loans made using alternative data and automated decisions are often small (e.g., to tide someone over until the end of the month), and so their results are possibly of limited utility. The new and growing market in automated lending using proprietary algorithms to evaluate borrowers with no traditional credit history is also highly innovative. Requiring new innovative lenders to share their lending results may deprive them of some of the benefits of their investment and first mover advantage. In addition, such firms are often entrepreneurial start-ups that may struggle with weighty reporting obligations as they seek to grow a risky business. Some do not even rely on credit reference bureau data themselves for their own lending decisions (relying entirely on alternative data), which may weaken the logic of reciprocity inherent in credit reference bureaus (where those supplying data are entitled to rely on the wider pool of aggregated data supplied by others).[112]

For these reasons, it is important to consider the overall data environment of the financial market as it develops, both in relation to the accuracy of data used in automated decisions and how responsibility for accurate data should be allocated in the formal credit data reporting systems and more generally.

Given the wide range of data available and its varying sources and levels of reliability, there are numerous policy dilemmas to come regarding how the guidelines on clarity and predictability in the fourth General Principle Credit Reporting ("The legal and regulatory framework should be sufficiently precise to allow service providers, data providers, users and data subjects to foresee consequences of their actions") will operate.

[111] See for example GPFI, *Use of Alternative Data* at footnote 29.

[112] For an excellent discussion of these issues, see Jason Blechman, *Mobile Credit in Kenya and Tanzania: Emerging Regulatory Challenges in Consumer Protection, Credit Reporting and Use of Customer Transactional Data*, African Journal of Information and Communication (AJIC), Issue 17, November 2016, http://www.macmillankeck.pro/publications.html.

Security, Infrastructure and Trust Working Group: *Big data, machine learning, consumer protection and privacy*

6.2 Bias and discriminatory treatment

Biased inferences and decision-making outputs

While one concern arising with big data is how input data, such as name, age and other personal data, will be used and protected, another relates to the inferences that result from processing such data. Just as important as the accuracy of the input data is the manner and accuracy of the inferences big data and machine learning will draw from it about individuals and groups, and the impact of such inferences on decisions.[113] Some such inferences, which predict future behaviour and are difficult to verify, may determine how individuals are viewed and evaluated and so affect their privacy, reputation and self-determination.

Data protection laws that govern the collection, use and sharing of personal data typically do not address the outputs of machine learning models that process such data. One of the concerns of data protection and privacy law and regulation is to prevent discrimination. Principle 5 of the High Level Principles for Digital Financial Inclusion states that data should "not be used in an unfair discriminatory manner in relation to digital financial services (e.g., to discriminate against women in relation to access to credit or insurance)."[114]

Recent examples of inferences involving major internet platforms concern sexual orientation, physical and mental health, pregnancy, race and political opinions. Such data may be used in decisions about whether a person is eligible for credit.[115] The GDPR sets apart special categories of personal data for tighter restrictions. While personal data is defined as "any information relating to an identified or identifiable natural person,"[116] "special categories" of personal data are more specific. They relate to "racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation."[117]

Limiting processing of special categories of data

Automated decision-making based on special categories of personal data is only permitted under the GDPR with explicit consent from the user or if "necessary for reasons of substantial public interest, on the basis of Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject."[118]

---

[113] Article 29 Data Protection Working Party, 'Opinion 03/2013 on Purpose Limitation, 00569/13/EN WP 203, Adopted on 2 April 2013' (2013) 47 http://ec.europa.eu/justice/article-29/documentation/opinionrecommendation/files/2013/wp203_en.pdf; Omer Tene and Jules Polonetsky, 'Big Data for All: Privacy and User Control in the Age of Analytics' (2012) 11 Nw. J. Tech. & Intell. Prop. https://scholarlycommons.law.northwestern.edu/cgi/viewcontent.cgi?article=1191&context=njtip; The European Data Protection Supervisor (EDPS), 'Opinion 3/2018 on Online Manipulation and Personal Data' 3/2018 8–16 https://edps.europa.eu/sites/edp/files/publication/18-03-19_online_manipulation_en.pdf.

[114] G-20, High Level Principles of Digital Financial Inclusion, p16, https://www.gpfi.org/sites/default/files/G20%20High%20Level%20Principles%20for%20Digital%20Financial%20Inclusion.pdf

[115] Astra Taylor and Jathan Sadowski, How Companies Turn Your Facebook Activity Into a Credit Score, The Nation, 15 June, 2015. https://www.thenation.com/article/how-companies-turn-your-facebook-activity-credit-score/

[116] GDPR, Article 4.

[117] GDPR, Article 9(4).

[118] GDPR, Article 22(4) and 9(2)(a) and (g).

The purpose of such tighter restrictions on dealing with special categories is to provide practical means of reinforcing other laws prohibiting discrimination on the basis of such data, whether in the provision of public or private services or otherwise. The right to privacy seeks to prevent disclosures that may lead to discrimination and other irreversible harms.[119]

In the era of big data, however, non-sensitive data can be used to infer sensitive data. For example, a name may be used to infer religion or place of birth which in turn can be used to infer race and other personal data that belong to the special categories. Shopping data can reveal purchase history of medicine from which a health condition may be inferred, affecting decisions such as a person's eligibility for health insurance.[120] Demographic and statistical data relating to wider groups may also be attributed to specific individuals. As a result, non-sensitive data may merit the same protections as sensitive data.[121] The result is that the distinction between sensitive and non-sensitive data becomes blurred and of questionable utility.[122]

This is not a light matter of definitional strain. One of the basic objectives of data protection and privacy law and regulation is to ensure that data is not used to result in discrimination, particularly of protected groups that have been the subject of historic discrimination. The nature of big data and machine learning undermines this objective. As several scholars put it recently, "A significant concern about automated decision making is that it could potentially systematize and conceal discrimination."[123]

Where machine learning algorithms are trained on input data that is based on historical examples, they may result in disadvantages for certain historically disadvantaged population groups. They may therefore reflect past discrimination regardless of the reasons that arose in the past (e.g., due to prejudice or implicit bias). Where such previous decisions were themselves biased, the training data for machine learning processes may perpetuate or exacerbate further bias.

An individual's creditworthiness may be evaluated based not only on their attributes, but those of their social network. In 2015, Facebook secured a patent that, among other things, enables filtering of loan applications depending on whether the average credit rating of a loan applicant's friends exceeds a prescribed minimum credit score.[124] This may risk discrimination, and even financial exclusion, if an applicant's friends are predominantly members of a low income population even if the applicant's own

---

[119] Article 29 Data Protection Working Party, 'Advice Paper on Special Categories of Data ("sensitive Data")' Ares(2011)444105-20/04/2011 10 available at http://ec.europa.eu/justice/data-protection/article-29/documentation/otherdocument/files/2011/2011_04_20_letter_artwp_mme_le_bail_directive_9546ec_annex1_en.pdf at p4.

[120] Antoinette Rouvroy, *"Of Data and Men": Fundamental Rights and Freedoms in a World of Big Data*, COUNCIL OF EUR., DIRECTORATE GEN. OF HUM. RTS. AND RULE OF L., at 10 (Jan. 11, 2016), https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016806a6020

[121] When considering automated decision-making, the Article 29 Working Party found that profiling can create sensitive data "by inference from other data which is not special category data in its own right but becomes so when combined with other data." Article 29 Data Protection Working Party, 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679', footnote 51, at 15.

[122] See Zarsky at footnote 15.

[123] See Joshua Kroll, Joanna Huey, Solon Barocas, Edward Felten, Joel Reidenberg, David Robinson, and Harlan Yu, *Accountable Algorithms*, Univ. of Penn Law Review, 2017, available at https://scholarship.law.upenn.edu/cgi/viewcontent.cgi?article=9570&context=penn_law_review. Paul Ohm and David Lehr, *Playing with the Data: What Legal Scholars Should Learn About Machine Learning*, Univ. of CA, Davis Law Review, 2017, available at https://lawreview.law.ucdavis.edu/issues/51/2/Symposium/51-2_Lehr_Ohm.pdf.

[124] http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO1&Sect2=HITOFF&d=PALL&p=1&u=%2Fnetahtml%2FPTO%2Fsrchnum.htm&r=1&f=G&l=50&s1=9100400.PN.&OS=PN/9100400&RS=PN/9100400

features should otherwise qualify him or her for the loan.[125] The risk is that, by relying on past data, such technologies will facilitate wealthier populations' access to financial services and impede access for minority groups that lacked access in the past, thereby "automating inequality."[126]

Discrimination may also be built into machine learning models in "feature selection," i.e., the choices in their construction regarding which data should be considered. While a model might not explicitly consider membership of a protected class (e.g., gender, race, religion, ethnicity), particularly if doing so would be unlawful, it might nevertheless rely on inputs that are effectively proxies for membership of such a protected class. Postcodes are a commonly cited example, as some areas have a high percentage of the population from a particular ethnic or racial group.

Another concern arises when the machine learning model fails to consider a wide enough set of factors to ensure that members of a protected group are assessed just as accurately as others. A model may have less credit data on members of a less advantaged group because fewer members of such group have borrowed in the past. If algorithms are trained using more input data from one particular group than another, they may produce outputs disproportionately inclined towards the former group.

Additionally, machine learning models could potentially be used to mask discrimination intentionally. This could arise if the training data is intentionally distorted or if proxies for a protected class are intentionally used in order to produce discriminatory results.

Techniques for removing bias based on a protected attribute focus on ensuring that an individual's predicted label is independent of their protected attributes.[127] However, even if protected attributes are not explicitly included, correlated attributes (proxies) may be included in the data set, resulting in outcomes that may be discriminatory. Addressing this in machine learning is challenging, but tests have been developed to assess the impact of an automated decision on different protected groups.[128]

In some countries, where bias is unintentional, it may nevertheless be unlawful if it has "disparate impact," which arises where the outcomes from a selection process are widely different for a protected class of persons (e.g., by gender, race or ethnicity or religion) compared with other groups despite the process appearing to be neutral. The notion of disparate impact was developed from a US Supreme Court decision in 1971[129] which found that certain intelligence test scores and high school diplomas were largely correlated with race to render discriminatory hiring decisions.[130] The legal theory was

---

[125] Jonathan Zim, *The Use of Social Data Raises Issues for Consumer Lending*, Miami Business Law Review, https://business-law-review.law.miami.edu/social-data-raises-issues-consumer-lending/.

[126] Virginia Eubanks, Automating Inequality, St Martin's Press (2018).

[127] Hardt, Moritz, Price, Eric, and Srebro, Nathan. *Equality of opportunity in supervised learning*, NIPS, 2017; Chouldechova, Alexandra, *Fair prediction with disparate impact: A study of bias in recidivism prediction instruments*, Corr, 2017.

[128] Disparate impact has been defined using the "80% rule" such that, where a dataset has protected attribute X (e.g., race, sex, religion, etc.) and a binary outcome to be predicted C (e.g., "will hire"), the dataset has disparate impact if:

$$\frac{\Pr(C = YES|X = 0)}{\Pr(C = YES|X = 1)} \leq \tau = 0.8$$

for positive outcome class YES and majority protected attribute 1 where Pr(C = cjX = x) denotes the conditional probability (evaluated over D) that the class outcome is c 2 C given protected attribute x 2 X. Feldman, Michael, Friedler, Sorelle A., Moeller, John, Scheidegger, Carlos, and Venkatasubramanian, Suresh, *Certifying and removing disparate impact*. In KDD, 2015. http://sorelle.friedler.net/papers/kdd_disparate_impact.pdf.

[129] Supreme Court of the United States. *Griggs v. Duke Power Co.* 401 U.S. 424, March 8, 1971.

[130] The US Supreme Court found that Duke Power's hiring decision was illegal if it resulted in "disparate impact" by race even though it was not explicitly determined based on race. This prevented Duke Power from using intelligence test scores and high school diplomas, qualifications largely correlated with race, to make hiring decisions. The legal doctrine of disparate impact that was developed from this

recently reaffirmed when in 2015 the US Supreme Court held that a plaintiff may establish a *prima facie* case against discrimination under the Fair Housing Act without evidence that it was intentional if they bring statistical proof that a governmental policy causes a disparate impact.[131]

The involvement of computers makes it more difficult to determine disparate impact, and thus bias. Disclosing and explaining the process of selection by algorithm may be difficult or effectively impossible. Nevertheless, where it can be shown that a model produces discriminatory results, it may be possible that it violates laws prohibiting discrimination, although proving this may be difficult, and justifications such as business necessity may also apply.[132]

Discriminatory selection could occur without involving protected groups. For instance, where digital financial services algorithms infer from user data that an individual is experiencing financial liquidity problems, payday lenders may be able to target vulnerable individuals with advertisements and offers for loans at high interest rates and charges. Competition from firms like ZestFinance may actually drive down the cost of lending to such groups, but concerns may arise if discriminatory selection has adverse results for an individual.[133]

Addressing discrimination tendencies

One approach to address machine learning's potential tendency towards discrimination is to incorporate randomness into the data.[134] For instance, a machine learning algorithm for extending credit may be trained using initial data that indicates that a certain group (e.g., from a particular postcode or of a particular gender or race) tends to have less reliable debtors. If the model were to extend credit to other groups, then a self-fulfilling prophecy may result whereby the characteristics of successful debtors correlate with non-membership of the protected group. Incorporating an element of randomness into the model so that some individuals who would not ordinarily be predicted to be reliable debtors nevertheless receive credit could allow the model to test the validity of the initial assumptions. The introduction of data that evolves to be closer to the real world may lead to improvements in the overall fairness and accuracy of the system.

Another suggested approach is to select or modify input data so that the output meets a fairness test operated by the system. Additional training samples from a minority group might be selected in order to avoid the model over-reflecting its minority status. There are other methods for ensuring statistical parity among groups that can be adopted,[135] and the important thing is to ensure that these are designed into the model, even using artificial intelligence to monitor artificial intelligence.

---

ruling is the main legal theory used to determine unintended discrimination in the USA. Duke Power was unable to prove that the intelligence tests or diploma requirements were relevant to the jobs for which they were hiring.

[131] *Texas Dep't of Housing and Community Affairs v. Inclusive Communities Project* 135 S. Ct. 2507 (2015)

[132] Solon Barocas and Andrew D. Selbst, *Big Data's Disparate Impact*, 104 Calif. L. Rev. 671 (2016). http://www.californialawreview.org/wp-content/uploads/2016/06/2Barocas-Selbst.pdf.

[133] Steve Lohr, *Big Data Underwriting for Payday Loans*, NY Times, January 19, 2015, https://bits.blogs.nytimes.com/2015/01/19/big-data-underwriting-for-payday-loans/.

[134] See *Accountable Algorithms*, at footnote 123.
Paul Ohm and David Lehr, *Playing with the Data: What Legal Scholars Should Learn About Machine Learning*, Univ. of CA, Davis Law Review, 2017, available at https://lawreview.law.ucdavis.edu/issues/51/2/Symposium/51-2_Lehr_Ohm.pdf.

[135] See *Accountable Algorithms*, at footnote 123.
Paul Ohm and David Lehr, *Playing with the Data: What Legal Scholars Should Learn About Machine Learning*, Univ. of CA, Davis Law Review, 2017, available at https://lawreview.law.ucdavis.edu/issues/51/2/Symposium/51-2_Lehr_Ohm.pdf.

In some cases, one might expect there to be a commercial incentive to remove bias. Bias is not only harmful to a service's reputation, but it may be suboptimal business economics for the service provider. If an applicant's postcode leads to a lower score and rejection of their loan application despite the applicant having a healthy income, low level of indebtedness and other positive attributes, then the lender has missed an opportunity to make a profitable loan.

In a perfect static market where providers compete on the same service and may refine it to increase market share, one might expect designers to improve algorithms over time to weed out bias. However, in a dynamic market where new models and services are constantly being developed with new data constantly being added, bias may be addressed only for the model to be updated or replaced by a new one that may reflect new bias, renewing the problem. Businesses may also focus more on rapid growth to win the new market, while viewing discriminatory impact on protected groups as a lower level priority. Even if the market might be expected over time to refine algorithms to reduce bias, in many cases it is simply socially and politically unacceptable to allow biases in the case of race, ethnicity and gender.

A key question is to what degree industry should bear the cost of identifying bias, using data to identify discrimination.

When automated decision-making causes unlawful discrimination and harm under existing laws, firms relying on such processing might employ tools (and, under some laws, they may be responsible) to ensure that using data will not amplify historical bias, and to use data processing methods that avoid using proxies for protected classes. In addition, human reviews of algorithm outputs may be necessary. It may also be possible to use data to identify discrimination, and to require companies by regulation to do so.

Even if the result may not violate existing laws prohibiting discrimination on the basis of race, religion or another protected class, the unfair harm to individuals may merit requiring industry to employ ethical frameworks and "best practices" to adjust algorithms to ensure that outcomes will be monitored and evaluated. Other mitigating measures may include providing individuals the opportunity (or right) to receive an explanation for automated decisions (see section 7.2), and employing data protection impact assessments (DPIAs) (see section 8).

Other approaches that have been suggested include consumer agencies randomly reviewing scoring systems of financial service providers (and health providers, educational institutions and other bodies that routinely make decisions about people) from time to time. They might run hypothetical scenarios to assess whether the models were effectively using statistical proxies for protected groups, such as race, gender, religion and disability. Such auditing might encourage firms to design against such risks.[136]

Differential pricing and other terms

Availability of data allows a financial service provider to better assess the risk that a consumer represents, and so to offer services that might not otherwise be available. However, the availability of a potentially vast array of data about a consumer also creates an information asymmetry whereby the provider knows more about the consumer than the consumer knows about the provider. The provider

[136] Danielle Keats Citron, *Technological Due Process*, Washington University Law Review, Vol. 85, pp. 1249-1313, 2007, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1012360

may take advantage of such situation and be able to engage in what economists refer to as "differential pricing," in which the provider charges different prices to different consumers for the same product.

Differential pricing is common and often has consumer benefits, for example, for train tickets are often sold at a discounted price to students and old age pensioners. It can, however, also result in perceived unfairness, where some population groups are targeted to pay higher prices based on their profile resulting from geographic location or other attributes.[137]

In financial services, the focus of differential pricing relates primarily to a consumer's risk profile. Pricing based on risk can improve economic efficiency by discouraging behaviour that is risky, rewarding individuals with no history of engaging in unlawful activities such as traffic accidents. It can improve access to insurance by reducing adverse selection, when only individuals with a high-risk profile will enrol at a uniform price. However, differential pricing of insurance products can result in unfairness where risk factors arise beyond an individual's control, e.g., in health insurance.

Big data may engage in differential pricing by drawing inferences from personal data about an individual's need for the service, and his or her capacity to pay and price sensitivity. The machine may estimate a price as near as possible to the maximum amount the profiled consumer may be willing to pay. Due to an asymmetry of information, the consumer does not know enough about the provider to negotiate the price down to the minimum amount the provider would be willing to accept (e.g., for it to achieve a reasonable return on investment).

In a dynamic market, competition would be expected to impose downward pressure on the provider's price, driving it downward towards its costs. However, policy concerns arise where differential pricing disadvantages persons who are already disadvantaged. An individual may be more desperate for a financial service, and thus be willing to pay a higher price. A lender may be able to charge a higher price that does not so much reflect the higher risk of default as the borrower's urgency. This may prejudice low income individuals and families.

Differential pricing can also become discriminatory where prices are set according to criteria that, while seemingly objective, result in adverse treatment of protected groups. For instance, if an algorithm sets higher prices for consumers with a postcode from a neighbourhood that has historically had higher levels of default than those from other neighbourhoods, individuals who do not themselves have other attributes to suggest a higher risk may face higher prices.

Certain historically disadvantaged population groups share particular attributes (such as a postcode). Individuals with those attributes may thereby suffer discrimination even if they do not have a bearing on creditworthiness. For example, a person with a healthy salary and little debt may be treated adversely as a result of living in a community (or having social media friends, or the same medical doctor, or shopped at discount stores) where people have historically higher debt-to-income ratios. Machine learning models are thus among other trends in automation of economic processes that may increase inequality over time.[138]

---

[137] Julia Angwin and Jeff Larson, *The Tiger Mom Tax: Asians Are Nearly Twice as Likely to Get a Higher Price from Princeton Review*, ProPublica, Sept. 1, 2015.

[138] See Karen Harris, Austin Kimson, and Andrew Schwedel, *Labor 2030: The Collision of Demographics, Automation and Inequality*, Bain & Company Report, February 7, 2018 available at http://www.bain.com/publications/articles/labor-2030-the-collision-of-demographics-automation-and-inequality.aspx.

6.3 Breach and re-identification

The vast amounts of data held by and transferred among big data players creates risks of data security breach, and thus risk to consumer privacy. Even when the amount of data held on an individual is kept to a minimum, their identity may be uncovered through reverse-engineering from even a small number of data points, risking violation of their privacy.[139] The risk of this occurring arises where the data may be obtained by third parties, whether through unauthorised access through a data breach or by transfer of the data to a third party with the agreement with the firm controlling or processing the data. In both cases, measures to protect the release of data about identifiable individuals include de-identification, pseudonymisation and anonymisation. Such measures and the challenges that they face in the context of big data are discussed in this section 6.3. Section 6.4 discusses the role and regulation of third-party intermediaries who acquire data by agreement in the data market.

The limits of de-identification, pseudonymisation and anonymisation

Personal privacy may be protected in varying degrees by using privacy enhancing technologies[140] (PETs) such as de-identification, which involves suppressing or adding noise to directly identifying and indirectly identifying information in a dataset, or otherwise introducing barriers (making it statistically unlikely) to identifying a person:[141]

- *Directly* identifying data identifies a person without additional information or by linking to information in the public domain (e.g., a person's name, telephone number, email address, photograph, social security number, or biometric identifiers).

- *Indirectly* identifying data includes attributes that can be used to identify a person, such as age, location and unique personal characteristics.

Whereas de-identification involves removing both of these, pseudonymisation removes only directly identifying data so that the personal data cannot be attributed to a specific individual without the use of additional information. Such additional information is kept separately and protected by technical and administrative measures to prevent such attribution.[142] The basic pseudonymisation process is not complex, simply substituting alternative attributes:

[139] See, e.g., Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1701, 1716-27 (2010).
[140] See FPF's Visual Guide to Practical Data.
[141] See Cavoukian, Ann and El-Emam, Khaled, De-Identification Protocols: Essential for Protecting Privacy, Information and Privacy Commissioner of Ontario, 2014; and Information Privacy Commissioner of Ontario, "De-Identification Centre", Information Privacy Commissioner of Ontario (https://www.ipc.on.ca/privacy/de-identification-centre/).
[142] GDPR, Article 4(5).

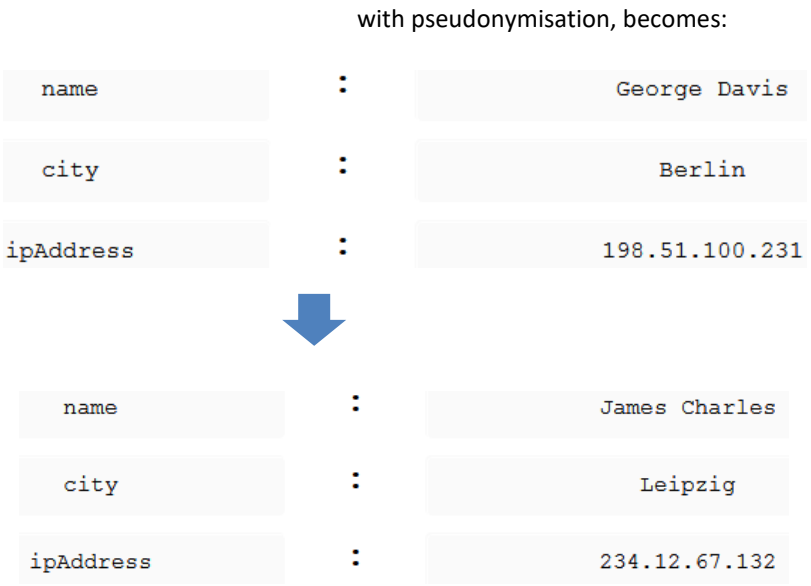

Figure 3 Pseudonymisation process. Source, KI Protect

De-identification is one means by which organisations can comply with "data minimisation" requirements in data protection laws, i.e., to collect, store and use only the personal data that is necessary and relevant for the purpose for which it is used (see section 5.1).

De-identification rarely eliminates the risk of re-identification. Re-identification may occur if de-identification was incorrectly implemented or controlled, or where it is possible to link de-identified data with already known personal data or publicly available information. Effective de-identification requires expert understanding of the data and the wider data ecosystem, including reasons and means by which adverse parties might seek to re-identify individuals.

Some experts criticise de-identification as being ineffective and as promoting a false sense of security by assuming unrealistic, artificially constrained models of what an adversary might do.[143] In a famous example in 1997, by linking health data that had been stripped of personal identifiers with publicly available voter registration data, it was possible to identify Governor William Weld of Massachusetts and thus link him to his medical records. (The Governor had previously assured constituents that their health data was kept confidential.)[144]

[143] Narayanan A, Felten EW. (Princeton). *No silver bullet: De-identification still doesn't work* 2014. Available at: http://randomwalker.info/publications/no-silver-bullet-de-identification.pdf.
[144] Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA LAW REVIEW 1701 (2010), https://www.uclalawreview.org/pdf/57-6-3.pdf.

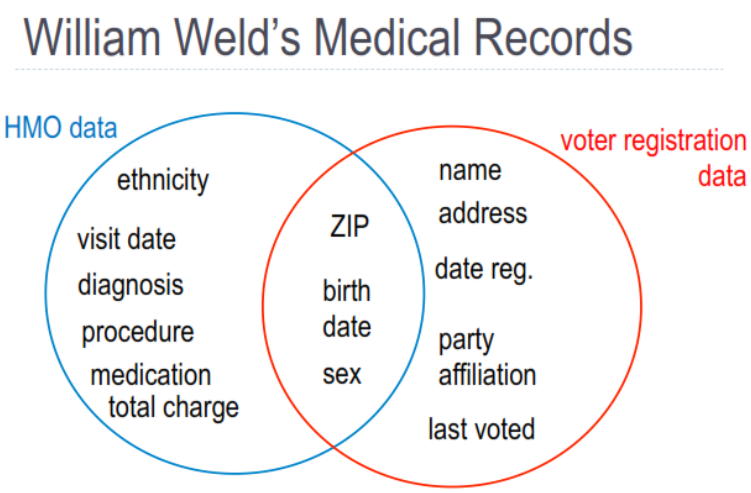

Figure 4. Identification of Governor Weld from four attributes

One study in 2013 found that 95% of mobility traces are uniquely identifiable given four random spatio-temporal points and over 50% of users are uniquely identifiable from two randomly chosen points (which will typically be home and work).[145] Richer data makes it possible to "name" an individual by a collection of fields or attributes, for example postal code, date of birth and sex.

Geolocation data carries particular risks of identification or re-identification of individuals. It is possible to combine user data linked to a persistent, non-unique identifier with other data to develop an enhanced profile of a person. Even geolocation data alone may be used to identify a user because the two most common user locations will typically be their home and work addresses. Sensitive data about an individual, for example a particular medical condition, may be identified due to their attendance at particular locations, such as an abortion clinic or mosque.

Measures may be employed to reduce such risks, such as accepting only insights rather than full datasets, accepting only data that has already been aggregated or de-identified, and applying additional filters where data is drawn from devices, e.g., accepting only geo-fenced data, removing home, work and sensitive locations or restricting the time of the data, and "blurring" or "fuzzing" datasets.

Anonymisation involves the elimination or transformation of the directly and indirectly identifying data. While pseudonymisation and de-identification involve procedures and technical, organisational and legal controls to prevent employees and third parties (such as researchers) from re-identifying individuals, anonymisation – once achieved – does not require such further measures. However, anonymisation reduces the utility of the data. The richer data is, the more useful it is.

Improving the approaches to re-identification risk

Technologies and criteria are emerging that seek to preserve the richness of data while reducing the identifiability of individuals. For instance, "differential privacy" has grown in popularity since Apple announced that it uses it to anonymise user data.[146] Differential privacy makes it possible to measure the quality of data anonymisation. It quantifies how much information the anonymisation method will leak about a given individual being added to a dataset using that method. It works with the trade-offs between utility and convenience, introducing random noise to eliminate the difference between what is revealed about an individual whose data is included in big data analysis and one who opts out.[147]

[145] Yves-Alexandre de Montjoye et al., *Unique in the Crowd: The privacy bounds of human mobility*, Scientific Reports 3 (2013).
[146] Apple, Differential Privacy Overview, https://images.apple.com/privacy/docs/Differential_Privacy_Overview.pdf.
[147] It is "a strong privacy guarantee for an individual's input to a (randomized) function or sequence of functions, which we call a privacy mechanism. Informally, the guarantee says that the behaviour of the mechanism is essentially unchanged independent of whether any individual opts into or opts out of the data set. Designed for statistical analysis, for example, of health or census data, the definition protects the privacy of individuals, and small groups of individuals, while permitting very different outcomes in the case of very different

Where the number of individuals involved is high enough, while the slightly biased statistical noise masks individuals' data, the noise averages out over large numbers of data points, allowing patterns to be detected and meaningful information to be emerge. This enables better discussion and decisions about trade-offs between privacy and statistical utility by providing a means of evaluating cumulative harm over multiple uses.

"[D]ifferentially private database mechanisms can make confidential data widely available for accurate data analysis, without resorting to data clean rooms, data usage agreements, data protection plans, or restricted views." Thus, it "addresses the paradox of learning nothing about an individual while learning useful information about a population."[148]

Statistical disclosure control, inference control, privacy-preserving data mining, and private data analysis are other algorithmic techniques that may be applied to large databases using statistical methods with a view to managing privacy.

A market is growing in services for de-identification, pseudonymisation and anonymisation. For instance, German company KIProtect[149] enables firms working with large datasets to secure the data, integrating over APIs with the client firm's data processing to detect and protect private or sensitive data by transforming the data using pseudonymization, anonymization and encryption techniques. The ability to support many data types and storage technologies (e.g., Apache Kafka and Google Firebase) allows use in a wide range of settings. The increasing availability of such service providers means that firms processing data can outsource key parts of their privacy needs, reducing the burden of building their own in-house privacy capability which is not their key business.

De-identification, pseudonymisation and anonymisation methodologies may not merely require to be included in the coding of dataset management, but also in administrative organisation. Thus, Apple performs differential privacy on user data on the user's device before Apple anonymises the user data (dropping IP addresses and other metadata) and collects, aggregates and analyses it. "Both the ingestion and aggregation stages are performed in a restricted access environment so even the privatized data isn't broadly accessible to Apple employees."[150]

In addition to these sorts of measures, a policy of "separation of duties" can reduce privacy risks in processing personal data. This limits any single administrator's power to a given role, with other roles managed by other administrators similarly limited, thus reducing the risk of a rogue administrator. Linked to this, a policy of "least privilege" would aim to ensure that each administrator will only have the powers necessary for their delegated function.

Ultimately, the difficulty of preventing re-identification may mean that a black-and-white view on de-identification may not be helpful, and the debate over the efficacy of these techniques may need to be looked at "in a more nuanced way, accepting that in some, but not all cases, de-identification might

data sets." Cynthia Dwork, *The differential privacy frontier*. In: Theory of Cryptography Conference, Springer, LNCS 5444. Berlin: Springer; 2009. pp. 496–502.

[148] Cynthia Dwork, *Differential Privacy*, 2006 PROC. 33RD INT'L COLLOQUIUM ON AUTOMATA, LANGUAGES & PROGRAMMING 1.

[149] See www.kiprotect.com.

[150] Differential Privacy Overview, at footnote 146.

provide acceptable answers."[151] Indeed, Cynthia Dwork suggests that continuous use of accurate data will eventually undermine privacy and the techniques mitigate rather than eliminate risk:[152]

> *[D]ata utility will eventually be consumed: the Fundamental Law of Information Recovery states that overly accurate answers to too many questions will destroy privacy in a spectacular way. The goal of algorithmic research on differential privacy is to postpone this inevitability as long as possible.*

In this light, regulation could seek to rely less on notification to consumers that their data will be collected, analysed and shared, and on obtaining their consent to this, and more on ensuring that privacy enhancing technologies are continuously integrated into big data and machine learning data processing and updated to deal with evolving challenges. Achieving this may require establishing incentives in legislation that create liability for data breaches, essentially placing less of the economic burden on the consumer by obtaining their consent and more on the organisations collecting, using and sharing the data.

6.4 Data intermediaries

Big data and machine learning are made possible not only by supply of data from online activity and demand from service providers that rely on it, but by intermediaries – the third-party data brokers who trade in personal data. This results in a huge number of sources of data, as well as methods of collection and data formats.

Various risks to the consumer arise with transfer of personal data. Transfer of data from one entity to another increases risk of breach due to the higher number of parties holding it, as well as from vulnerabilities of the transfer process itself. Sensitive, confidential data may be obtained by third parties without permission, risking identity theft, intrusive marketing and other privacy violations.

The very transfer of data to a third party may itself be something that the consumer might not have expected when originally sharing their data with a company, for example when accessing its service or when merely browsing the internet. Lastly, the proliferation of data about a person may increase the asymmetry of bargaining power between consumers and the firms selling them products and services, as discussed in section 4.2.

The transfer of data from one entity to another means that an organisation processing the data will often have no direct relationship with the original entity that collected it, and indeed, it may be at several levels of remove. The acquiring entity may lack information about whether the data was collected and is transferred in compliance with data protection and privacy laws.

Where data is obtained with user consent (e.g., credit card use data, financial transaction data, email data), the key question will be whether consent was validly obtained. For data obtained from public spaces (e.g., satellite insights data, drone data, surveillance footage, dropcam data), the key question will be whether the data was really obtained from public spaces, and in a manner consistent with surveillance laws. Where data was obtained from the internet without express user consent (web scraping, documented and undocumented APIs), the issue will be whether the data was obtained

[151] Professor Pompeu Casanovas, Louis De Koker, Danuta Mendelson, Professor David Watts, *Regulation of Big Data: Perspectives on Strategy, Policy, Law and Privacy*, Health and Technology (2017), available at: https://ssrn.com/abstract=2989689.
[152] *Differential Privacy*, at footnote 148.

through authorized access. Certification approaches may emerge whereby data may be guaranteed to have been subject to de-identification, pseudonymisation and anonymisation before it is traded.

Currently the market in data is very fluid. Firms buy and sell data, and reduce their risk of liability and thus economic burden associated with data privacy, by obtaining contractual representations and warranties about compliance with privacy laws, such as whether any necessary user consent was obtained. Companies such as ZwillGen[153] will advise firms relying on big data how to manage their economic risks arising from privacy law liability.

Little of this provides reassurance to the individual subject of the data. It also raises questions about the responsibility of entities that acquire data downstream, including in relation to the levels of due diligence they should perform. The difficulty of tracking data processing and transfer operations adds complexity to the problem of attributing responsibility for the unauthorised use of personal data.

Data brokers are, therefore, coming under increasing scrutiny, including providing consumers direct rights. For instance, the US FTC singled out data brokers in its Privacy Report to allow consumers to access their data through an easy-to-find, easy-to-use common portal, and supported legislation that would allow consumers to access, and a right to dispute or suppress, data held by brokers.[154]

In May 2018, the small US State of Vermont was the first to enact *An Act relating to data brokers and consumer protection*.[155] This new law regulates businesses that collect, sell or license to third parties' personal information of Vermont residents with whom the business does not have a direct relationship. It requires data brokers to register as such with the authorities, disclose information about their data collection activities, and maintain security measures to protect the data. Failure to do so is a violation of Vermont's consumer protection laws, which may lead to enforcement by the Attorney General or by a private citizen. California's new Consumer Privacy Act 2018 also imposes restrictions on transfers of data to brokers.[156]

The development of laws governing data brokers promises to open up a new area of consumer rights to access data held on them, rectify incorrect data, and obtain redress for violations of their rights.

7 Post-engagement: accountability

When complex automated decision-making systems operate without human involvement, there is a need to ensure that creators, designers, manufacturers, operators, maintainers, and users of the algorithms and systems are accountable for their respective elements in the process. Achieving this requires transparency or traceability, whereby the automated decision maker can explain the decision and its rationale for rejecting other possible decisions in favour of the one chosen. This requires documentation of each decision made about the data selected, its treatment and the design of algorithms. Lastly, the creators, designers, manufacturers, operators, maintainers, and users of the algorithms and systems must bear economic responsibility for their decisions, where appropriate in the form of legal liability.

---

[153] https://www.zwillgen.com/.
[154] US Federal Trade Commission, Protecting Consumer Privacy in an Era of Rapid Change: Recommendations for Businesses and Policymakers 27 (Mar. 2012), http://www.ftc.gov/os/2012/03/120326privacyreport.pdf.
[155] An act relating to data brokers and consumer protection, House Bill 764 ("H-764"), available *here*.
[156] See https://iapp.org/resources/article/california-consumer-privacy-act-of-2018/ and https://iapp.org/resources/topics/california-consumer-privacy-act/.

This section considers consumers' rights where something has gone wrong after they have shared data or after personal data about them (shared by them or by others) has otherwise been used to their disadvantage or harm. It begins in section 7.1 by reviewing consumer rights to address problems with data that could be used in decisions affecting them. This includes rights to rectify incorrect data held about them and to have certain data erased. This is discussed here in this post-engagement section because it arises after a firm has obtained an individual's personal data, but it may of course merely be a prelude to another engagement when the data will be used.

This section then considers the consumer's position after personal data about them has been used in big data and machine learning in a way that affects them, such as a decision with legal or similar consequences. It considers difficulties big data and machine learning pose to traditional approaches to transparency and accountability in section 7.2, including the problem of a right to obtain an explanation of inferences and decisions based on them. Section 7.3 then reviews the consumer's rights to contest decisions that have been made about them using big data and machine learning processes. Lastly, the question of showing that harm has actually been suffered is discussed in section 7.4. Accountability cannot work without liability on the backend.

7.1 Rights of access, rectification and erasure

A key safeguard for consumers in data protection and privacy laws is the right to access data held by an organisation about the individual and to rectify errors in it, or complete it if it is incomplete.[157]

For instance, the recently enacted California Consumer Privacy Act of 2018 requires businesses that collect personal information of California residents, if a consumer requests, to disclose (without charging) the types of personal information it has collected about that consumer over the previous year. This includes the specific pieces of information collected and categories of third parties with which the information has been shared.[158] The EU's GDPR confers rights on individuals to be informed if personal data about them is being processed, to receive a free copy of that data,[159] to have inaccuracies corrected, and to complete personal data that is incomplete.[160]

Such rights are also widely recognised in international law.[161] The OECD Privacy Handbook says, "[t]he right of individuals to access and challenge personal data is generally regarded as perhaps the most important privacy protection safeguard."

---

[157] According to the *OECD Privacy Handbook*, "[t]he right of individuals to access and challenge personal data is generally regarded as perhaps the most important privacy protection safeguard". *OECD Privacy Handbook*, 2013, Chapter 3 (Explanatory Memorandum for Original 1980 Guidelines). In Europe, see *Case C-131/12, Google Spain v. Agencia de Protección de Datos* (AEPD), 2014 EUR-Lex (May 13, 2014). See Kelly & Satola, "The Right to Be Forgotten", University of Illinois Law Review, Vol. 1, 2017.

[158] California Consumer Privacy Act of 2018, CAL. COV. CODE §§178.110(a) & (b), 178.130(a)(2).

[159] 2016 EU General Data Protection Regulation, Article 15.

[160] *Ibid*, Article 16.

[161] The General Comment 16 on Article 17 of the *International Covenant on Civil and Political Rights* provides that "every individual should have the right to request rectification or elimination" of files containing incorrect personal data. Human Rights Committee, General Comment 16 (on Article 17 on the right to privacy), 1988, Supp. No. 40, UN Doc A/43/40, para 10. A General Comment to an international convention is a non-binding guide to its interpretation. The *Council of Europe Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (Convention 108)* provides for "rectification or erasure" of any data processed contrary to the principles on data quality, which require that personal data undergoing processing must be adequate and up-to-date. Council of Europe Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (1981), Art 8(c), with reference to Art 5. Similarly, under the *APEC Privacy Framework*, individuals should have the right to "challenge the accuracy of information relating to them and, if possible an as appropriate, have the information rectified, completed, amended or deleted". 2004 APEC Privacy Framework, Art. 23(c).

In some jurisdictions, the individual may have the right to access not merely provided data and observed data, but also inferred data and derived data (see section 4.3). These may include profiles that the data controller has developed, and information about the purpose of the data processing, the categories of data held and their source.[162]

Rectification may be simple for a consumer where the data is verifiable, such as their date of birth, address, salary level or marital status. However, in the case of big data and machine learning, data about the individual may comprise inferences rather than the plain facts of their life.

Some inferences, such as a person's predicted levels of income, expenses or illnesses over time, or age of death, may be important to automated (or human) decisions about an individual, such as for example eligibility for, or price of, financial services. Some suggest that individuals' rights to rectify data ought not to be restricted to verifiable personal data because the verifiability of an inference may not determine its effect on the individual concerned, and because the individual may be able to provide information that supplements the inference (e.g., updated health information).[163]

An increasing number of data protection laws provide individuals with the right of erasure (also referred to as the right to be forgotten) of personal data about them where the data are no longer necessary for the purposes for which they were collected or processed.[164] Under the GDPR, individuals have the right to erasure of personal data about them where the data are no longer necessary in relation to the purposes for which they were collected or processed and, if the processing is based on consent, where the individual withdraws that consent and there is no other legal ground for the processing.[165] The right to be forgotten was famously exercised in Spain against Google.[166] California's new law also requires businesses to comply with a consumer's request to delete personal information unless the information is necessary for the business to perform certain functions.[167]

Whether inferences drawn through machine learning may be the subject of a right of access, rectification or erasure has not as yet been established, and in many countries is not certain. It is likely that most countries' data protection laws will be applied to give greater weight to the interest of a business in retaining and using data it has produced through machine learning processing, than the privacy interests of consumers, just as its trade secrets and intellectual property will be attributed value

[162] Article 29 Data Protection Working Party, 'Guidelines on the Right to Data Portability' (2017) 16/EN WP 242 rev.01 10, available at https://ec.europa.eu/newsroom/document.cfm?doc_id=44099.
[163] Article 29 Data Protection Working Party, 'Opinion 4/2007 on the Concept of Personal Data; 01248/07/EN WP 136' (n 68) 6; Article 29 Data Protection Working Party, 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679' (n 19) 18.
[164] GDPR, Article 17. See Kelly & Satola, "The Right to Be Forgotten", University of Illinois Law Review, Vol. 1, 2017. California's new Consumer Privacy Act requires certain businesses to meet a consumer's request to delete personal information unless the information is necessary for the business to perform certain functions. California Consumer Privacy Act of 2018, CAL. COV. CODE, §178.105.
[165] GDPR, Article 17.
[166] *Case C-131/12, Google Spain v. Agencia de Protección de Datos* (AEPD), 2014 EUR-Lex (May 13, 2014). A Spanish national complained to the Spanish Data Protection Agency (AEPD) about Internet stories linking his name with attachment proceedings in a real-estate auction related to recovery of social security debts. Mr Costeja González requested that the newspaper remove or alter the pages, or that Google Spain or Google Inc remove or conceal the personal data in search results. Google objected to the Spanish National High Court, which requested a decision of the European Court of Justice (ECJ), which found that Google was a data controller against which the right to be forgotten could be exercised, and thus Mr. Costeja had the right to make the request and have it reviewed by the AEPD. See Kelly & Satola, *The Right to Be Forgotten*, University of Illinois Law Review, Vol. 1, 2017.
[167] *Ibid*, §178.105.

compared with the consumer's potentially nebulous interests.[168] Of course, data may already have been shared with third parties before the consumer requests its erasure, further weakening this remedy.

In a big data era, the proliferation of personal data about individuals poses important challenges to individuals' ability to exercise these rights.

7.2 Transparency and explanations

Explaining automated decisions made using big data and machine learning

Accountability for decisions typically begins with or at least requires an explanation for the basis and method of the decision.[169]

Some advocate establishing (as some jurisdictions such as the EU have done) a consumer right to an explanation where a solely automated decision, such as a declined loan application or reduction in a credit limit, has legal or other significant effects.[170]

However, two problems arise in providing an explanation to the consumer in the context of big data and machine learning:

First, the techniques are hard to explain, particularly in plain language to consumers. Machine learning models are described as "opaque"[171] and as "black boxes."[172] Even providing source code will not inform even the computer scientists how a decision was made, as "[m]achine learning is the science of getting computers to act without being explicitly programmed."[173]

Second, to some degree, the machine learning models are the subject of trade secrets and software copyright that are the result of investment and exist in a competitive commercial market. A machine learning operator may be reluctant to share the coding of or an explanation for the machine learning algorithm lest this weaken competitive opportunity and undermine the initial investment.

These factors present important challenges for accountability to consumers for the use of algorithms.[174] In particular, the difficulty of explaining to a consumer the relationship between data inputs and outputs is a barrier to the consumer challenging decisions made about them. Nevertheless, even if explanations are currently difficult to generate, it may be that only if such legal rights are created will the necessary efforts be made.

There may be important reasons to make such efforts. Society-wide acceptance of big data and machine learning, particularly automated decision-making and the services that rely on it, will depend at least in part on trust – trust that the relevant information has been considered in a reasonable manner. It is a

---

[168] See, e.g., Gianclaudio Malgieri, 'Trade Secrets v Personal Data: A Possible Solution for Balancing Rights' (2016) 6 International Data Privacy Law 102, 115.
[169] Finale Doshi-Velez and others, 'Accountability of AI Under the Law: The Role of Explanation' [2017] arXiv preprint arXiv:1711.01134.
[170] See *Ethically Aligned Design*, at footnote 217 at p160.
[171] Jenna Burrell, 'How the Machine "Thinks:" Understanding Opacity in Machine Learning Algorithms' [2016] Big Data & Society.
[172] See Frank Pasquale, The Black Box Society: The Secret Algorithms That Control Money and Information, Harvard University Press (2015).
[173] Stanford Univ., Machine Learning, COURSERA, https://www.coursera.org/learn/machine-learning/home/info [https://perma.cc/L7KF-CDY4]
[174] See *Accountable Algorithms*, at footnote 123.
Paul Ohm and David Lehr, *Playing with the Data: What Legal Scholars Should Learn About Machine Learning*, Univ. of CA, Davis Law Review, 2017, available at https://lawreview.law.ucdavis.edu/issues/51/2/Symposium/51-2_Lehr_Ohm.pdf.

common perception that in machine learning, correlation and prediction are the governing principles, and that causality and reasoning are unimportant. In 2008, Chris Anderson declared the scientific method obsolete, overtaken by the corroborative power of mass correlations.[175] Machine learning identifies correlations between factors, which do not amount to causation. It may be able to make predictions for future behaviour, but not explain the reasons.

Machine learning occurs where a computer system is exposed to large quantities of data (from historical examples), is trained to observe patterns in the data, and infers a rule from those patterns. Rather than establishing rules directly, humans generate a computerised rule-making process. This abstraction, or disconnect, between the humans and the decision, creates challenges for verifying the rules that are created. This makes it difficult to hold them accountable when the rules or their results fail to meet policy goals, or even fall foul of laws, particularly relating to discrimination. Indeed, not only do ordinary people not understand machine learning models, but even those who develop them are often unable to explain why they succeed.

However, in many sectors, it is not workable for machine learning models to be understood only by data scientists and computer programmers. In medicine, banking, insurance and other sectors, researchers and even practitioners must understand the machine learning models they rely on if they are to trust them and their results. Trade-offs may arise between keeping models and modelling processes transparent and interpretable (which requires minimising complexity) and developing machine learning models that evolve over time to improve their accuracy and performance (which makes them more complex and harder to explain).

Furthermore, the accuracy of machine learning depends on how data used for training and validation of machine learning models is selected and curated. It also depends on articulating properly the task of the model, allowing for well-developed hypotheses, and selecting relevant metrics for performance. Ultimately, given enough time and resources, a computer programme should be explainable, or otherwise there can be no reason to have confidence in the accuracy of its conclusions.[176]

While some suggest that complexity defies explanation, others argue that such a view conceals the ready understandability of algorithms, and that "rather than discounting systems which cause bad outcomes as fundamentally inscrutable and therefore uncontrollable, we should simply label the application of inadequate technology what it is: malpractice, committed by a system's controller."[177] Still, there are clearly challenges to providing explanations for automated decisions that can be readily understood by inexpert humans.

Regulating for adequate explanations

When a financial service provider makes a decision based on data inputs (e.g., income and asset levels, post code), the decision is ultimately based on inferences made from these sources, such as whether the individual's risk of default on a loan of a certain size over a certain period is too high to justify the loan. Typically, data protection laws do not provide protection against unreasonable inferences, leaving such matters to sector specific laws, if at all. Indeed, most data protection laws do not require the data

[175] Chris Anderson, The End of Theory: The Data Deluge Makes the Scientific Method Obsolete, Wired, 23 June 2008, https://www.wired.com/2008/06/pb-theory/

[176] Hildebrandt, Mireille, Preregistration of machine learning research design. Against P-hacking in: BEING PROFILED:COGITAS ERGO SUM, ed. Emre Bayamlıoğlu, Irina Baraliuc , Liisa Janssens, Mireille Hildebrandt Amsterdam University Press 2018 (forthcoming) (September 27, 2018). Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3256146

[177] Kroll JA. 2018 *The fallacy of inscrutability.* Phil. Trans. R. Soc. A 376, 20180084. (doi:10.1098/rsta.2018.0084)

controller to provide an explanation for an automated decision that has been made. At most, they typically require notifying a person that a future decision will be automated, and perhaps offer an opportunity to opt out of it.[178]

Some countries go a little further. For instance, Brazil's Data Protection Act 2018 provides the consumer with the right to request a review of decisions taken solely on the basis of automated processing of personal data affecting their interests. This includes decisions designed to define his profile or evaluate aspects of his personality, and the right to request clear and relevant information on the criteria and procedures used for the automated decision.[179]

Some policy makers do lean towards greater scrutiny of automated decisions under data protection and privacy law. The EU's Article 29 Data Protection Working Party, for instance, advised that data controllers should avoid over-reliance on correlations, and should provide meaningful information to the concerned individual about the logic involved in automated decision-making.[180] Such disclosures might include the main characteristics considered in reaching the decision, the source of this information and its relevance. In the same vein, data controllers may be required to show that their models are reliable by verifying their statistical accuracy and correct inaccuracies, particularly to prevent discriminatory decisions.[181]

The Future of Privacy Forum has suggested that explaining machine learning models should include documenting how the model was chosen, providing a legal and technical analysis to support this. This would include identifying the trade-offs between explainability and accuracy. It would record decisions to make a model more complex despite the impact of diminished explainability, and take account of the materiality of the output to individuals and third parties (e.g., there is more at stake in medical treatment than movie recommendations).[182]

Some argue that the lack of effective explanations presents an accountability gap, and that data protection and privacy laws should confer on consumers an effective "right to reasonable inferences."[183]

Where inferences carry high risk of rendering adverse decisions, harming reputation or invading privacy, such a right could require a data controller to explain before processing (*ex ante*) the relevance of certain data for the inferences to be drawn, the relevance of the inferences for the type of automated decision and processing, and the accuracy and statistical reliability of the method used. Such explanations could be supported by an opportunity to challenge decisions after they are made (*ex post*).

This would permit, in addition to contesting an automated decision on the basis of accuracy of its inputs, challenging verifiable inferences on which it is based, such as the individual's level of income or assets, health, or relationship status. Non-verifiable inferences might be challenged by provision of supplemental data that might alter their conclusions.

---

[178] Sandra Wachter, Brent Mittelstadt and Luciano Floridi, 'Why There Is No Right to Explanation in the General Data Protection Regulation' [2017] International Data Privacy Law https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2903469.
[179] Article 22.
[180] GDPR, Articles 13-15.
[181] Article 29 Data Protection Working Party, 'Guidelines on Automated Individual Decision Making and Profiling for the Purposes of Regulation 2016/679', see footnote 51, at p28-29.
[182] Future of Privacy Forum, Beyond Explainability: A Practical Guide to Managing Risk in Machine Learning Models (2018).
[183] See Wachter & Mittelstadt, at footnote 52.

Efforts to introduce regulation that intrudes into the substance of decisions or the process of decision-making, as opposed to the mere collection, use and sharing of data, may be viewed by some as burdening a nascent innovative sector that should be left to develop products that benefit consumers, and refine them under competitive pressure. Others will view it as seeking to rebalance the disempowerment of consumers resulting from the removal of human elements in key stages of decision-making (see further in section 7.3). In a human interaction, the individual may have an opportunity to meet or speak with a decision-maker or someone who can influence the decision-maker, and to explain where inferences were erroneous. For the right to human intervention in automated decisions to have substance, it may require fleshing out the ultimate integrity of the process that the human intervention aspires to achieve.

Data protection laws do not typically guarantee the accuracy of decision-making, and this likely generally extends to the accuracy of inference data, so that even where incorrect inferences have been drawn from accurate data, the individual may not have a right to rectify such inferences.[184]

This would more typically be the remit of sector-specific laws, such as a financial services law, but in most countries, such laws will only prohibit decision-making that is discriminatory according to specified criteria (such as race, gender or religion) and not prescribe the correctness of the decision itself. In this sense, a poor algorithm is similar to a poor bank clerk who fails to make a good decision due to poor judgment or inexperience: it may be poor business practice but is not unlawful.

However, a financial services law may proscribe certain procedures intended to ensure that decisions are more likely to be good ones. For instance, it may require a financial service provider to carry out an assessment of the customer's need that will make it more likely that a product suits him or her.[185] It could also require risk assessments that will ensure that risks are considered, including in the algorithms themselves.

Improving explanations

An alternative or supplement to providing an explanation has also been suggested – that consumers should be provided "counterfactual" feedback on automated (and only predominantly automated) decisions, positive or negative. Counterfactual explanations can inform the concerned individual not so much how a decision was reached but rather what variations in the input data might have led to a different decision.[186] For instance, a digital financial service provider could inform the consumer, "Your loan application stated that your annual income is $30,000. If your income were $45,000, you would have been offered a loan."[187]

---

[184] See Wachter & Mittelstadt, at footnote 52.

[185] E.g., Central Bank of Kenya, Guideline on Consumer Protection, Section 3.2.1(c)(iv), requires that banks not: take advantage of a consumer who is not able to understand the character or nature of a proposed transaction. [A bank] shall therefore inquire of the consumer's specific needs and shall provide suitable products or services relevant to those needs. While Section 3.2.2(i) of the Guideline states "Depending on the nature of the transaction and based on information provided by a customer, [a bank] should assess and understand the needs of the customer before rendering a service." In addition, Section 3.2.4(a)(ii) also requires banks, when giving advice to customers, ensure that "any product or service which the institution recommends to a consumer to buy is suitable for the consumer."

[186] Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Harvard University Press 2015); Viktor Mayer-Schönberger and Thomas Ramge, *Reinventing Capitalism in the Age of Big Data* (Basic Books 2018).Sandra Wachter, Brent Mittelstadt and Chris Russell, 'Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR' [2017] arXiv preprint arXiv:1711.00399; *Accountable Algorithms*, at footnote 123.

[187] Sandra Wachter, Brent Mittelstadt & Chris Russell, *Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR*, Harvard Journal of Law & Technology, 2018. https://arxiv.org/abs/1711.00399

Of course, there are many input variables to decision making, and many combinations of such variables that could produce a near infinite number of potential counterfactuals. Thus, it is unlikely that one can reduce an explanation for a decision to one or even a few variables. In addition, such an approach would need to be wary of the commitment it may make to offer the service on the alternative terms (if the individual then presents with an income of $45,000, they might have a legitimate expectation that the loan will be approved).

However, if such counterfactuals were coded into the service, the counterfactual results could be provided rapidly to the consumer, who could potentially experiment with different levels of variables. Indeed, consumer interfaces could even provide a sliding scale for inputs, allowing some experimentation by the consumer. It may thus be possible to provide some counterfactuals that would improve the consumer's understanding, and offer an opportunity to contest the decision, or even to modify their situation to allow a more favourable decision. For instance, by understanding that stopping smoking would entitle the individual to health insurance, or that paying off a certain debt or increasing his or her income would result in a positive credit decision, the individual can exercise more affirmative agency over his or her life than being the passive recipient of the decision.

This may narrow the gap in negotiating positions and result in a commercially profitable offer for a desired service to be made and accepted, benefitting both provider and consumer. There may, then, be reasons to expect market participants to introduce such features as a differentiating element of their services in a competitive market, although a regulatory "nudge" could be useful in some cases to get such practices started and make them mainstream.

It has been suggested that the counterfactual approach might also mitigate concerns that requiring explanations may lead to exposure of trade secrets and violations of non-disclosure obligations. Providing counterfactuals may avoid having to disclose the internal logic of the algorithms of the decision-making system. This could be a practical, results-oriented approach to transparency, and may have advantages over requirements to provide an explanation that may be so complex that it neither increases understanding nor enables improvements in a consumer's situation.

While referring to counterfactuals is a relatively light means of improving the position of consumers, not least in opening up alternative means to obtain the services they seek, there are deeper ways to improve accountability of machine learning systems. It might be possible, for instance, to review and certify properties of computer systems, and to ensure that automated decisions are reached in accordance with rules that have been agreed upon, for example to protect against discrimination. Such an approach is referred to by some as "procedural regularity." [188]

For a machine learning model to function in an accountable manner, accountability must be designed into the system. System designers, and those who oversee design need to begin with accountability and oversight in mind. The IEEE's Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems recommends:[189]

> *Although it is acknowledged this cannot be done currently, A/IS should be designed so that they always are able, when asked, to show the registered process which led to their actions to their*

[188] See *Accountable Algorithms*, at footnote 123.
Paul Ohm and David Lehr, *Playing with the Data: What Legal Scholars Should Learn About Machine Learning*, Univ. of CA, Davis Law Review, 2017, available at https://lawreview.law.ucdavis.edu/issues/51/2/Symposium/51-2_Lehr_Ohm.pdf.
[189] See Ethically Aligned Design, at footnote 217 at p159.

> *human user, identify to the extent possible sources of uncertainty, and state any assumptions relied upon.*

The IEEE also proposes designing and programming AI systems "with transparency and accountability as primary objectives,"[190] and to "proactively inform users" of their uncertainty.[191]

7.3 Right to contest decisions

As discussed in section 7.2, data protection laws typically do not give a right to contest the accuracy of decisions made with their data. However, consumers are increasingly provided the opportunity to contest decisions made on the basis solely of automated processing. Novel risks arise from automated decision-making in life-affecting areas of financial services such as credit, insurance and risky or costly financial products.[192] The IEEE Global Initiative recommends that "Individuals should be provided a forum to make a case for extenuating circumstances that the AI system may not appreciate—in other words, a recourse to a human appeal."[193] An increasing number of data protection and privacy laws, including the GDPR, provide the right to obtain human intervention, express one's views and contest the decision.[194]

Such a right originates from notions of due process, which may be undermined if decisions are made by a machine without further recourse. It also originates from the view that treating people with respect and dignity includes ensuring that important decisions over their lives involve not merely a machine but another human being. This concern is amplified by the risk of machines producing erroneous results or behaving discriminatorily.[195]

The ability to contest an automated decision is not merely a matter of clicking a request for reconsideration and receiving another, final automated decision, which would then just produce another automated decision subject to a right to contest it. Ultimately, if an automated decision is to be reviewed, it would be necessary to ensure that the automated decision is subject to some form of human intervention, where the individual has an opportunity to present their point of view to another human being who will consider whether the automated decision should be revised.

Such human intervention may vary in its degree of involvement, from a full right of appeal of the entire substance of the matter, to merely a check that the algorithm did at least receive accurate data inputs without verifying its functionality. Overall, however, it is likely that such rights to contest decisions with human intervention will be limited to cases where the input data was incorrect or incomplete, the requisite consent of the individual was not obtained, or there was some other infringement of data protection principles. One might describe these as more procedural than substantive matters. The "reasoning" behind the substance of decisions, which inhabits the design and functioning of algorithms, would likely not be subject to contest under data protection laws.

This does not mean that sector-specific laws, regulations and standards cannot require providers to modify or nullify their decisions where they are generated by machine learning models for substantive

[190] *Ibid* at p152.
[191] *Ibid* at p159.
[192] Article 29 Data Protection Working Party, 'Guidelines on Automated Individual Decision Making and Profiling for the Purposes of Regulation 2016/679', see footnote 51, at p10.
[193] IEEE Global Initiative (see footnote 217) at p153.
[194] GDPR, Article 22(3).
[195] Andrea Roth,*Trial by Machine*, 104 GEO. L.J. 1245 (2016).

reasons. However, it does mean that until such laws, regulations or standards are introduced, consumers have limited recourse to challenge an automated decision.[196]

While individuals may be protected from prescribed collection, use and sharing of their personal data (particularly sensitive or special categories of data) and the accuracy and completeness of their data used in automated decisions about them, they have little protection when it comes to the way decisions are actually made.

7.4 Harm and liability

Accountability depends ultimately on being held responsible in law, including compensating for harm that has been caused. One difficulty of developing policy, legal obligations and remedies for consumers in the area of data protection arises from the intangible nature of the harm against which the consumer requires to be protected, or for which they need to be compensated.

This can undermine a consumer's claim from the get-go. To have standing in a court to bring a claim to recover compensation, it is typically necessary to allege that one has been harmed. Courts have struggled to identify harm from data protection and privacy law violations, often producing very different legal views. Many claims have been dismissed because consumers failed to show the harm they have suffered.

Whether or not a person has suffered harm is often considered against a counterfactual, i.e., whether the person is put in a worse position than if the event had not happened.[197] Demonstrating harm is particularly challenging where there has not yet been any pecuniary or physical loss, for instance where a system has been breached and data has been obtained without permission but it has not (yet) been used to steal money. Harm may be viewed as conjectural, whereas in some legal systems, plaintiffs must show that they have in fact suffered injury.[198]

Theories of harm from personal data being obtained unlawfully include risk of fraud or identity theft, and anxiety the individual may experience about such risks. While intangible injuries are more difficult to recognise and analyse, they can be just as real and concrete as pecuniary damage.[199] Indeed, not only may intangible harms be genuine, it is increasingly argued that the very *risk* of harm – i.e., where damage has not yet materialised but the risk is present – should be treated as legitimate harm for the purpose of consumer claims.

Such harm may be evaluated according to the likelihood and magnitude of future injury, the sensitivity of data exposed, the possibility of mitigating harms and the reasonableness of preventative measures.[200] Courts have tended to be more sympathetic to plaintiffs in the case of identity theft due to risk of fraud,[201] or where inaccurate information about a person is published.[202]

---

[196] Wachter & Mittelstadt, footnote 52.
[197] Joel Feinberg, *Wrongful Life and the Counterfactual Element in Harming*, in FREEDOM AND FULFILLMENT 3 (1992).
[198] For example, the US Supreme Court in *Clapper v. Amnesty International* 133 S. Ct. 1138 (2013) rejected claims against the US Government for increased collection of data for surveillance reasons on the basis that the plaintiffs had not shown "injury in fact."
[199] *Spokeo*, ibid.
[200] Daniel J. Solove and Danielle Keats Citron, *Risk and Anxiety: A Theory of Data Breach Harms*, 96 Texas Law Review 737 (2018).
[201] In *Remijas v. Neiman Marcus Group*, LLC, 794 F.3d 688, 693-94 (7th Cir. 2015), the US Federal Court found that the fact that plaintiffs knew that their personal credit card information had been stolen by individuals who planned to misuse it (as other plaintiffs' cards had been the subject of fraudulent misuse) was sufficient harm to give them standing to sue.
[202] In *Spokeo* (see footnote 109), the US Supreme Court found that when a people search engine described a person incorrectly, this could potentially cause enough risk of harm to allow him standing to sue.

In the case of automated decision-making, there are various potential types of harm.[203] These may impose economic loss on a person, for example through denying, or raising the price of goods or services due to a person's classification as a member of a particular group (e.g., a person's neighbourhood, sometimes called "redlining"). A person may suffer a loss of opportunity, for example as a result of filtering candidates for a loan, credit limit increase or insurance contract according to race, genetic or health information.

Some harms are unlawful in some countries where they involve discrimination on the basis of race, religion, criminal history or health. In these cases, existing laws will specifically protect certain classes of people and may prohibit discriminatory outcomes. However, where membership of a protected class is not involved, there may be little way to show harm.

Another difficulty facing consumers harmed by big data and machine learning systems is identifying who should be held liable for the damage – for example, the firm employing the system, the firm that coded the algorithms, the firm that supplied the data? Demonstrating the precise cause and tracing the responsible party may be impossible for the consumer.

Section 6.2 discussed various things that operators of machine learning systems can do to reduce risk of bias. In addition to these, some have suggested requiring some firms relying on artificial intelligence and machine learning to obtain insurance, or other guarantees of financial responsibility, to provide a means of redress for those harmed.[204] While this may be more immediately obvious for personal injury cases involving equipment such as autonomous vehicles than claims for lost opportunity, it might be considered for cases of harm caused by data breaches by processors of large data sets.

It has also been suggested that when courts and legislators address claims for some form of injury resulting from artificial intelligence and machine learning, they should draw from the rich body of product liability law. This might in some cases mean applying strict liability, i.e., without showing causation, negligence or fault (let alone intention), for certain harms. Again, redress mechanisms should incentivise providers to address the problems both before and after they arise. For example, product liability law often seeks to avoid undermining the incentive of manufacturers to fix faults after their products cause harm out of fear that this will be treated as an admission of responsibility for the harm. In such cases, the law will provide that such steps are not admissible as evidence of fault.[205]

Overall, much remains to be done in most jurisdictions to give consumers effective remedies for breaches of their privacy and risks of big data and machine learning.

8 Risk management, design and ethics

The previous sections have discussed consumer protection and data privacy, focusing on legal and regulatory treatment and remedies. The resulting uncertainty presents a risk to business of being held responsible for violating antidiscrimination laws or incurring substantial liability for damages for privacy violations and data security breaches. This section looks at various steps that companies can take to mitigate these risks.

[203] For a lively description of these, see Cathy O'Neill, Weapons of Math Destruction (2016). For a useful taxonomy of potential harms from automated decision-making, see Future of Privacy Forum, *Unfairness by Algorithm: Distilling the Harms of Automated Decision-Making*, December 2017.
[204] IEEE Global Initiative (see footnote 217) at p156.
[205] IEEE Global Initiative (see footnote 217) at p156.

Security, Infrastructure and Trust Working Group: *Big data, machine learning, consumer protection and privacy*

8.1 Risk management

A common approach in situations of uncertainty is to apply risk management frameworks and processes, and thus good big data model design includes building risk management into the model.[206] For example, some financial service providers like Mastercard will apply the cross-industry process for data mining (CRISP/DM), which provides a structured approach to planning data mining projects.[207]

Such frameworks and processes may be employed to assess risks associated with consumer privacy and discrimination, just as any other risk. The US National Institute of Standards and Technology (NIST) recently launched work on a Privacy Framework,[208] focusing on risk management approaches modelled on its Cyber Security Framework. This framework emphasises the importance of prioritising risk management over "tick-the-box" compliance approaches.

Risk management processes for machine learning systems might include documenting objectives and assumptions, and employing "three lines of defence" that ensure separation (by process, roles, parties involved and incentives) of:

- development and testing of a machine learning model;
- its validation and legal review; and
- periodic auditing of the model throughout its lifecycle. [209]

Ongoing monitoring, improvement and accountability of machine learning systems depends on documenting these objectives.[210]

Risk management may apply to both input and output data in machine learning models:[211]

On the ***input data*** side, risk mitigation will start with documenting the requirements of the model (e.g., data-freshness, features and uses), the degree of dependence on data from surrounding systems, why and how personal data is included and how it is protected (e.g., encryption or otherwise), as well as its traceability. Such documentation supports effective review and maintenance. It will include assessing the "completeness, accuracy, consistency, timeliness, duplication, validity, availability, and provenance" of the input data. Mechanisms to ensure the model may be tested, updated and monitored over time may also be important.

On the ***output data*** side, various processes may be instituted to reduce risk of machine learning models producing adverse results. Bias detection mechanisms can be instilled to ensure that population groups are not discriminated against, or at least bias is quantified and minimised. Sometimes it may be necessary to restrict certain types of data in the model. Output data can also be analysed to detect proxies for features that might be a basis for discrimination, such as gender, race or postal code. This

---

[206] *Ibid.*
[207] See https://www.sv-europe.com/crisp-dm-methodology/.
[208] See https://www.nist.gov/privacy-framework.
[209] See for example Guidance on Model Risk Management, Board of Governors of the Federal Reserve System & Office of the Comptroller of the Currency, April 2011, available at https://www.federalreserve.gov/supervisionreg/srletters/sr1107a1.pdf; and the European frameworks, Directive 2013/36/EU of the European Parliament and of the Council of 26 June 2013 on access to the activity of credit institutions and the prudential supervision of credit institutions and investment firms; Regulation No. 575/2013 of the European Parliament and of the Council of 26 June 2013 on prudential requirements for credit institutions and investment firms; and the European Central Bank guide to the Targeted Review of Internal Models (the TRIM Guide).
[210] Thus the IEEE's Global Initiative (see footnote 217) recommends that "Automated systems should generate audit trails recording the facts and law supporting decisions."
[211] The following summary of risk management in machine learning is drawn from Future of Privacy Forum, Beyond Explainability: A Practical Guide to Managing Risk in Machine Learning Models (2018).

requires guidance from lawyers regarding the types of features that would be an unlawful basis for discrimination. Constant monitoring through statistical representation of output data should also improve detection of anomalies, feedback loops and other misbehaviour. Again, documenting these and ongoing testing will improve and widen understanding of a model's risks.

Risk assessment extends both to the input and output data, and to the creation and operation of algorithms. The research institute AINow[212] has proposed that public agencies carry out "algorithmic impact assessments", including in procurement of data and software, and in the operation of automated decision-making processes, as part of a wider set of accountability measures.[213]

Altogether, data processors need to define intended outcomes as well as unintended outcomes that should be avoided (working with legal and compliance teams), and be ready to correct or pull the model out of usage. If outputs risk breaching consumer protection, data privacy, antidiscrimination or other laws, firms should be ready with a strategy for dealing with authorities. For instance, California's guidance on permits for autonomous vehicles has specific provisions addressing how a firm should interact with law enforcement if there is an accident or another unintended outcome.

Part of the correct functioning of algorithms, including to prevent future harm, involves ensuring continued maintenance. Some have called for an ongoing legal requirement to monitor outcomes from algorithms, provide mechanisms for receiving feedback (e.g., complaints), conduct inspections, and correct models. [214] Such sophisticated matters are beyond the capability of consumers, who lack expertise and resources. Sometimes human monitoring will be important, not merely as part of an appeal from a consumer, but as part of the decision-making process itself. Such human involvement needs to be thoughtfully explored.

8.2 Integrating data privacy by design

Effectively addressing consumer protection and data privacy in big data and machine learning will require going beyond laws and regulations, and tick-the-box compliance with them. It will need to include designing products and services to minimise invasion of privacy. The seven principles of privacy by design developed under the leadership of Ann Cavoukian[215] are:

1. Be proactive not reactive, preventative not remedial, anticipating and preventing privacy-invasive events before they happen;
2. Make privacy the default setting so that consumers do not have to change settings to protect privacy, i.e., use opt-in rather than opt-out consents;
3. Embed privacy into design, integral to the system without diminishing functionality as opposed to bolted on after design (e.g., including the feature of data portability;
4. Adopt a win-win approach, benefitting from stronger consumer trust, lower risk from data breach;
5. Employ end-to-end security, ensuring secure intake, storage and destruction of data over the life cycle (including encryption of data storage and transfer);

[212] https://ainowinstitute.org/.
[213] Dillon Reisman, Jason Schultz, Kate Crawford, Meredith Whittaker, *Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability*, April 2018, https://ainowinstitute.org/aiareport2018.pdf.
[214] IEEE Global Initiative (see footnote 217) at p156.
[215] Ann Cavoukian, *Privacy by Design: The Seven Foundational Principles* (Information and Privacy Commissioner of Ontario, 2011), https://www.ipc.on.ca/wp-content/uploads/Resources/7foundationalprinciples.pdf; and David Medine, *Privacy by Design for Financial Services* https://www.livemint.com/Opinion/1ShpKAOC59VlXiwgCkVv8O/Privacy-by-design-for-financial-services.html.

6. Show visibility and transparency, using policies and keeping records to enable internal monitoring and independent verification; and
7. Demonstrate respect for user privacy, providing individuals access to information and the opportunity to contest and correct, complete and update data about them.

It will require privacy engineering in product development, including integration into training of computer scientists. For instance, Carnegie Mellon University offers a Masters of Science in Information Technology – Privacy Engineering program that addresses a range of such subjects.[216]

8.3 Ethics

Beyond management and engineering, there are broader efforts underway to change underlying attitudes and awareness of those in the tech industry. These include steps in the engineering community to develop ethics for artificial intelligence and autonomous decision-making. Bodies such as the Association for Computing Machinery[217] and the Institute of Electrical and Electronic Engineers are examples.[218] These measures alone do not secure fairness, accountability and transparency, but they do provide a vocabulary and value system that enables far more rapid communication about these topics, and make it far easier to develop the necessary risk management, engineering and other measures that lead to greater protection for consumer privacy.

[216] The Carnegie Melon programme includes: 1) Design cutting-edge products and services that leverage big data while preserving privacy; 2) Propose and evaluate solutions to mitigate privacy risks; 3) Understand how privacy-enhancing technologies can be used to reduce privacy risks; 4) Use techniques to aggregate and de-identify data, and understand the limits of de-identification; 5) Understand current privacy regulatory and self-regulatory frameworks; 6) Understand current technology-related privacy issues; 7) Conduct privacy-related risk assessments and compliance reviews, respond to incidents, and integrate privacy into the software engineering lifecycle phases; 8) Conduct basic usability evaluations to assess the usability user acceptance of privacy-related features and processes; and 9) Serve as an effective privacy subject-matter expert, working with interdisciplinary teams. Masters of Science in Information Technology – Privacy Engineering program. See also https://bigid.com/the-advent-of-privacy-engineering/.

[217] When computers decide: European Recommendations on Machine-Learned Automated Decision Making, Informatics Europe & EUACM 2018. http://www.informatics-europe.org/news/435-ethics_adm.html

[218] IEEE, Ethically Aligned Design: A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems version 2 (2018), https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf. A previous version 1 was published for consultation in 2016, http://standards.ieee.org/develop/indconn/ec/ead_v1.pdf.

9 Areas for further exploration

This paper has explored various challenges that consumer protection and data privacy law and regulation face with regard to big data and machine learning techniques, particularly where these are used for making decisions about services provided to consumers. Conventional requirements to provide notice of the intended purpose of using a consumer's personal data when the purpose may as yet be unclear, or obtaining consent for something the consumer largely cannot understand, are under strain. Risks from inaccuracy of data inputs, or bias and discriminatory treatment in machine learning decisions also raise difficult questions about how to ensure that consumers are not unfairly treated. The difficulty of ensuring transparency over decisions generated by algorithms, or of showing what harm has been caused by artificial intelligence techniques that would not have otherwise been caused, also pose challenges for consumer protection and data privacy law and regulation.

There are various areas where further work can be usefully advanced to develop standards that can apply across big data and machine learning, to work towards a balance between freedom to innovate and protection of consumers and their data privacy. These might include:

1. Developing standards for integrating ***privacy principles in the design*** of artificial intelligence and machine learning models. Following the principles developed by Ann Cavoukian (see section 8.2), these might include standards for (1) proactive design approach, (2) use of privacy default settings, (3) adoption of privacy by design, (4) consumer-trust orientation, (5) end-to-end security, (6) consumer access to information and the opportunity to contest and correct, complete and update data about them, as well as (7) standards for generating, recording and reporting logs and audit trails of the design process to enable review, and ensuring that such logs and audit trails are coded into the system.

2. Developing ***ethical standards*** for artificial intelligence computer programming to which the community of developers may refer to address the sorts of issues discussed in this paper, and which may be the basis of ongoing discussion for identifying new issues and how to approach them.

3. Developing standards for ***acceptable inferential analytics***. These could address assessment of output data and decisions of machine learning models against privacy and antidiscrimination principles. They could also address when inferences of personal attributes (e.g., political opinions, sexual orientation or health) from different sources of data (e.g., internet browsing) are acceptable or privacy-invasive depending on the context. This might also include developing standards for establishing the reliability of inferences, particularly those with high social importance, risk and legal effect, and in relation to protected groups. In addition, standards could be developed for testing inferences before and after deployment. Such standards may require different approaches to different types of services.

4. Developing standards for ***explanations*** of automated decisions, including asserting the relevance of data used to inferences drawn by the system, the relevance of such inferences for the type of automated decision, and the accuracy and statistical reliability of the data and methods used. This could involve encouraging developers of scoring models to share with consumers (and if required, regulators) the key attributes used in a model, and their relative weighting, and ensuring that

documentation and audit trails are provided in case of legal process. Developing standards for explanations could also include examining the potential for using counterfactuals to inform the consumer how, with different input attributes, they might obtain different decisions from the automated decision-making system. In circumstances where these are considered to be viable, standards could be developed for providing post-decision counterfactual explanations.

5. Developing best practices in processes for allowing consumers to obtain ***human intervention***, as well as for identifying the appropriate degree of human intervention that maintains the integrity and value of the model, while also offering the consumer a meaningful opportunity to be heard by a human being.

6. Developing principles of international best practice and harmonisation of ***accountability mechanisms***, including procedures for contesting automated decisions, standards for establishing *prima facie* harm, and ultimately frameworks for assessing liability for design and operation of artificial intelligence and machine learning models.