Building a high-performing data ethics programme from the ground up

Received (in revised form): 11th November, 2022

Alexandra Ross

Senior Director, Autodesk, USA

Alexandra Ross is Senior Director, Senior Data Protection, Use & Ethics Counsel at Autodesk, Inc. where she provides legal, strategic and governance support for Autodesk's global privacy, security, data use and ethics programmes. She is also an adviser to BreachRx and an Innovators Evangelist for The Rise of Privacy Tech (TROPT). Previously she was Senior Counsel at Paragon Legal and Associate General Counsel for Wal-Mart Stores. She is a certified information privacy professional (CIPP/US, CIPP/E, CIPM, CIPT, FIP and PLS) and a frequent speaker at privacy and security conferences. She holds a law degree from Hastings College of Law and a BS in theatre from Northwestern University. Alexandra is a recipient of the 2019 Bay Area Corporate Counsel Award — Privacy. Alexandra launched The Privacy Guru blog in 2014 and has published an ebook *Privacy for Humans* (available on Amazon and iTunes).

E-mail: alexandra.ross@autodesk.com

Ilana Golbin

Director, PwC, USA

Ilana Golbin is a Director — Emerging Technology & AI, Global Responsible AI Leader in PwC Labs, where she serves as one of the leads for artificial intelligence. Ilana specialises in applying machine learning and simulation modelling. She is listed as one of 100 'Brilliant Women in AI Ethics' in 2020, was recently recognised in Forbes as one of 15 leaders advancing ethical AI and is an active contributor to working groups and standards efforts for AI governance. Since 2018, she has led PwC's efforts globally in the development of cutting-edge approaches to build and deploy responsible AI.

E-mail: ilana.a.golbin@pwc.com

Bret S. Cohen

Partner, Hogan Lovells, USA

Bret Cohen helps technology and brick-and-mortar businesses comply with privacy, cyber security, Internet and consumer protection laws. He also represents companies in litigation and government investigations in these areas. As a lawyer and technologist, Bret has a knack for translating legal standards into practical technical requirements that are easy for clients to use. With a particular focus on the Internet and e-commerce, Bret has advised extensively on legal issues related to cloud computing, social media, mobile applications, online tracking, and analytics and software development. He counsels and is a frequent speaker on strategic compliance with global privacy laws, including cross-border transfer restrictions, data localisation requirements and the impact of government surveillance on the digital economy. Bret also spearheads efforts on cyber security incident preparedness and response, student privacy, marketing privacy and workplace privacy. With the global privacy landscape rapidly changing, Bret stays plugged into policymaking developments worldwide to help companies focus on both the legal requirements of today and the likely obligations of the future. He writes regularly on these developments, including for the Hogan Lovells Chronicle of Data Protection blog, for which he serves as the Managing Editor. Bret also strongly believes that privacy is not a zero-sum proposition and advises companies on how adopting privacy-enhancing policies and technologies can benefit business and enhance customer engagement in the long run. During law school, Bret was an articles editor for *The George Washington Law Review* and was awarded the Laurence E. Seibel Memorial Award for Excellence in Labor and Employment Law. He graduated with honours from the University of Maryland with dual degrees in computer science and business information systems.

E-mail: bret.cohen@hoganlovells.com

Abstract Companies are presented with increasingly complex legal, ethical and operational challenges when implementing algorithmic data processing to detect security threats or generate business insights. In this paper we will share leading practices on how to navigate the compliance landscape, build and maintain an ethics-by-design programme for data and technology, leverage existing frameworks and manage stakeholders. We will also introduce emerging technical concepts from the privacy and security domains and provide a perspective of how these technical concepts may be introduced into the governance process for organisations.

KEYWORDS: governance, artificial intelligence, machine learning, data ethics, data protection

INTRODUCTION

Algorithms are increasingly used in both the private and public sector to make or inform important decisions that once were made solely by humans. In particular, companies' reliance on artificial intelligence and machine learning (AI/ML) is on the rise, given the power of these technologies to identify correlations within unstructured data sets more quickly than ever before.

But the benefits of AI/ML do not come without risks. The predictive processes used by AI/ML can lead to harm, or to disparate harms across different groups of people, if not appropriately overseen or monitored. Consider, for example, the use of facial recognition technology by government agencies and law enforcement. The Government Accounting Office (GAO) found in 2021 that 20 out of 42 federal agencies used facial recognition technology,[1] and Georgetown estimated in 2016 that at least a quarter of state and local police departments could run searches in facial recognition systems.[2] Yet, in a seminal study, researchers found that facial recognition algorithms exhibited both skin-type and gender bias, resulting in classification being up to 20 per cent less accurate for darker-skinned faces relative to lighter-skinned faces and for female faces compared to male faces.[3] If agencies use facial recognition algorithms that are potentially biased, they run the risk of misidentifying suspects across racial or gendered lines. This is not merely hypothetical. For example, one article documented the cases of three Black men who were falsely arrested in 2020 based on incorrect facial recognition matches.[4]

This potentially disparate impact of AI is demonstrated in many other critical contexts. For example, independent researchers found that a Massachusetts hospital's clinical algorithm underestimated kidney disease risk for Black patients, which led to Black patients being placed relatively lower on donation recipient lists than others.[5] In the education context, during the COVID-19 pandemic, the UK replaced student exam results with algorithmically standardised predictions of exam results. One input in the algorithm included student rankings, which was weighted more heavily for students at smaller schools and effectively afforded them with grade inflation. This algorithmic process ended up giving lower scores to students living in traditionally lower-income neighbourhoods.[6]

These examples of biased outcomes, among many others, have led to broad public concern about the misuse of AI/ML. In turn, this concern has translated into the passage of many new laws and regulations aiming to prevent or mitigate harms and the application of existing laws, such as anti-discrimination laws, to these use cases. As use of these types of algorithms increases, it will become crucial for organisations to build governance programmes that address both ethical considerations as well as legal obligations arising from their use.

UNDERSTANDING DATA ETHICS

Recognising the risk of harm and potential liabilities of AI/ML tools, many organisations have started to establish governance structures to assess ethical considerations in the development, implementation and use of algorithms and other technology used to evaluate data sets and predict outcomes.

But defining 'data ethics' — and incorporating it into an organisation's decision-making processes — is not entirely straightforward. It requires careful thought about an organisation's values and principles. As defined by the Open Data Institute, data ethics is 'a branch of ethics that evaluates data practices with the potential to adversely impact on people and society — in data collection, sharing and use'. Alternatively, the U.S. Federal Data Strategy has defined data ethics as 'the norms of behavior that promote appropriate judgments and accountability when acquiring, managing, or using data, with the goals of protecting civil liberties, minimizing risks to individuals and society, and maximizing the public good'.

In short, data ethics focuses on the individual and societal harms posed by the collection, analysis, and use of analysis of large data sets. These ethical problems can arise for different reasons, including unrepresentative data sets,[7] flawed nature of prediction[8] and incorrect correlations.[9]

As a result, the use of predictive algorithms such as AI/ML can lead to concerns regarding privacy, profiling, automated decision making, civil rights and data governance. Data ethics encourages companies to look beyond the valuable products, services and innovations they bring to the market, to evaluate their outputs and downstream effects. It forces organisations to ask: do the end results of our uses of technology harm people or cause disproportionate harm to certain groups of people?

EVOLVING LEGAL FRAMEWORKS

The ethical uncertainties inherent in AI/ML technologies present policymakers with challenging questions

Governments around the world have increasingly enacted new laws, and developed new interpretations of existing laws, to govern the use of predictive algorithms derived from large data sets. Many of these laws impose a duty on organisations to set up governance and accountability structures designed to minimise the harms arising from the use of algorithms, requiring organisations to regularly review their technology and systems for potential instances of discrimination or unfair outcomes. At the same time, while there are similar concepts in play, a look at recent legal developments in the European Union (EU) and US shows a variety of approaches within and between jurisdictions as policymakers grapple with these new challenges.

European Union

Perhaps the most prominent example of an existing law in this space is the EU's General Data Protection Regulation (GDPR), which incorporates requirements that directly affect the use of decision-making algorithms. Notably, article 22 of the GDPR provides individuals with a default right not to be subject to a decision based solely on the algorithmic processing of their personal data, if that decision provides legally significant effects for that individual, such as the denial of credit or an employment opportunity.[10] Separately, the GDPR requires organisations to build data protection by default and by design into the development and use of algorithms analysing personal data,[11] to provide appropriate notice to consumers about the use of automated decision-making systems (including 'meaningful information about the logic involved')[12] and to undertake data protection impact assessments (DPIA) for uses of data that have the potential to significantly harm individuals.[13]

The newly adopted EU Digital Services Act (DSA) governs online platforms, aiming to hold them accountable for potential harms arising from illegal content, harmful advertising and disinformation that consumers suffer due to use of those platforms. One of the DSA's key features is the grant of strong enforcement powers to European authorities to monitor and investigate online platforms, including their algorithm use and data practices. Companies are required to provide access to, and explanations relating to, databases and algorithms during on-site inspections and other investigations.

There is also a pending AI Regulation in the EU that focuses on data governance and would establish different governance and risk management requirements based on the level of risk posed by an AI system.[14] It would require companies, prior to enacting certain applications of AI/ML, to analyse data sets that are used in the training, validation and testing of ML, including identifying potential biases, checking for inaccuracies and assessing suitability of the data. It also includes transparency requirements during the development phase of high-risk AI systems, robust post-market monitoring and evaluation of AI systems, and reporting and investigations of any AI-related incidents or errors.

United States

Since 2018, five US states have passed comprehensive privacy laws that provide consumers with varying combinations of rights that have implications for algorithmic processing. A commonly established consumer right in US state privacy laws is the right to opt out of certain types of data collection and use, such as profiling. The California Consumer Privacy Act (CCPA) defines profiling as the automated processing of personal information that is used to evaluate, analyse or predict aspects of an individual's work performance, economic situation, health, personal preferences, interests, reliability, behavior, location or movements.[15] These state privacy laws also tend to prohibit the use of automated decision making for 'legal or similarly significant effects' and to require data protection assessments for technologies that appear to demonstrate a high risk of profiling or a reasonably foreseeable risk of harm.

In addition, the US Federal Trade Commission (FTC) has highlighted the importance of accounting for ethical and legal considerations when designing and using predictive algorithms, as well as the potential enforcement consequences companies may face if they do not consider these factors. In 2021, the FTC issued guidance warning companies that the sale or use of biased algorithms may constitute an unfair or deceptive trade practice that violates Section 5 of the FTC Act, emphasising that companies should be transparent about their use of algorithms and their impacts, be truthful when making claims about algorithmic capabilities and vet algorithms for their impact on people before launch.[16] The FTC also flagged the existence of anti-discrimination laws such as the Fair Credit Reporting Act (FCRA) and the Equal Credit Opportunities Act (ECOA), which impose requirements when making credit, employment or insurance eligibility determinations, and indicated that biased outcomes could trigger an enforcement action under these acts, even if due in part or in whole to an automated process. Notably, the FTC emphasised that they would use the wide range of tools at their disposal to regulate these uses of algorithms. In some cases where the FTC has determined that some companies have been unlawfully using data to develop AI/ML models, the FTC has not only ordered the companies to delete the data itself but also to delete the models and algorithms developed using that data.

Rising public concern over this issue has also led to a consistent stream of proposed legislation at the federal level focused on identifying and mitigating algorithmic bias and discrimination, with evocative names such as the Protecting Americans from Dangerous Algorithms Act[17] and the Justice Against Malicious Algorithms Act.[18] In 2022, even the first bipartisan, bicameral federal privacy bill, the American Data Privacy and Protection Act, included a section requiring companies to conduct algorithmic impact assessments, with particular focus on potential harms related to individuals under age 17; decisions related to housing, education, employment, healthcare, insurance or credit opportunities; and disparate impact based on race, colour, religion, national origin, sex or disability status.[19] There have also been bills and resolutions about artificial intelligence introduced in over 20 states, many focused on anti-discrimination principles, notice, risk assessment and reporting.

GOVERNANCE AND RISK MANAGEMENT FOR AI

The proliferation of AI/ML and algorithm-centred regulations

The proliferation of AI/ML and algorithm-centred regulations in the EU, US and around the globe reflects a growing public concern that AI will be used in ways that could disadvantage some people over others and that companies are not well-positioned to prevent those harms or mitigate those risks. Organisations will likely find themselves increasingly obliged — whether based on law or public opinion — to implement ethical considerations into their AI systems to help ensure their technology does not result in harm. Effective governance structures are an important tool that organisations can use to incorporate data ethics and manage the related risks.

Applying existing governance models to AI

The Institute of Internal Auditors (IIA) defines governance as structures and processes that enable accountability and actions, supported by independent advice.[20] This approach, supported by organisations such as the IIA, centres around a Three Lines of Defence model, which has been adopted by many organisations with robust internal audit and governance functions. In this model, the first line is responsible for delivering services. The second line is responsible for defining standards and practices. The third line, independent of the other two, is responsible for providing advice around the performance of the first and second lines. All three lines report to a leadership team (see Figure 1).

The first line has several major functions:

1. Business units that serve customers directly;
2. Internal services that create or manage capabilities to support the business;
3. Product teams that build capabilities the business can sell.

In today's world, this first line is further split by technical expertise, around centres of excellence for the delivery of algorithmic systems, analytics and automations like robotic process automation (RPA). At the same time, this technical first line — the builders — is increasingly interested in participating in and owning portions of the governance programme, which can add great value. Expanding governance programmes to include these teams may be necessary to capture the risks posed by algorithmic processing of data and the use and deployment of AI, as well as other emerging technologies.

This technical group approaches governance with the goal of mitigating risks at the point of development, and they are increasingly interested in leveraging technology-driven solutions to do so. There is no shortage of vendors[21] purporting to provide 'ethical' technology platforms,

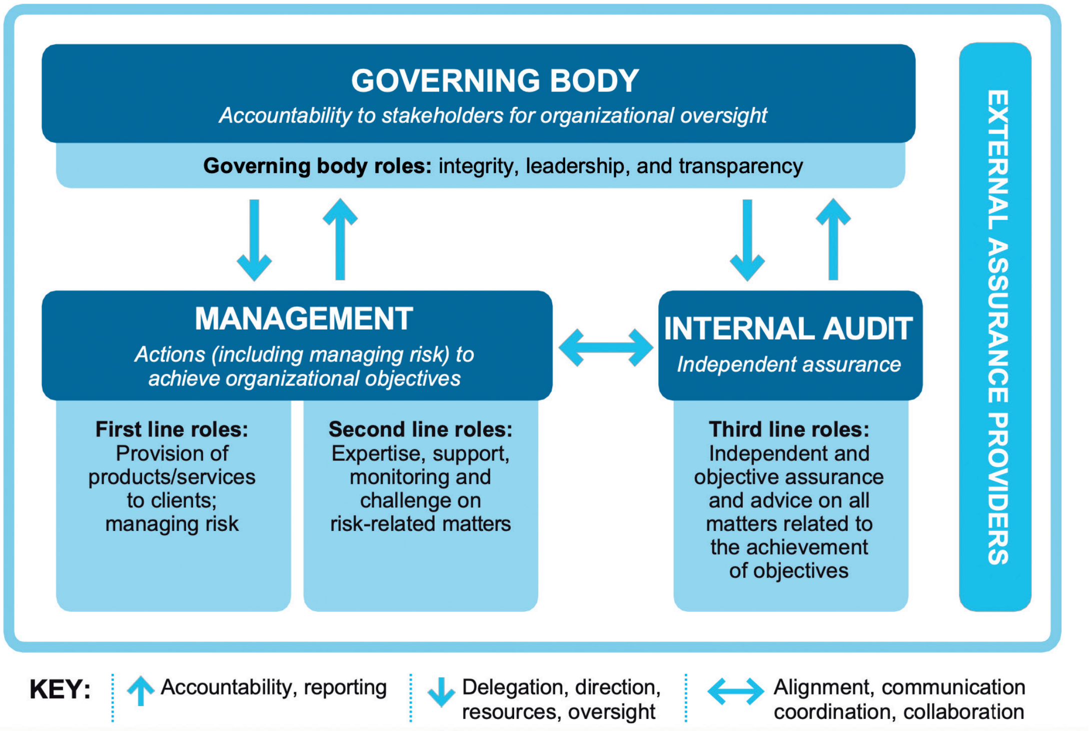

GOVERNING BODY
Accountability to stakeholders for organizational oversight
Governing body roles: integrity, leadership, and transparency
MANAGEMENT
Actions (including managing risk) to achieve organizational objectives
INTERNAL AUDIT
Independent assurance
First line roles: Provision of products/services to clients; managing risk
Second line roles: Expertise, support, monitoring and challenge on risk-related matters
Third line roles: Independent and objective assurance and advice on all matters related to the achievement of objectives
EXTERNAL ASSURANCE PROVIDERS
KEY:
Accountability, reporting
Delegation, direction, resources, oversight
Alignment, communication coordination, collaboration

Figure 1: IIA Three Lines of Defence Model
Source: IIA

and the temptation to use tools to solve a technology-exacerbated problem is great.

Additionally, the open-source community has come together around several major research initiatives to improve our use of data in algorithmic systems: bias mitigation, privacy, security, ML operations and governance being just a few of these topic areas. Many of these problem areas are highly interrelated; for example, a bias risk could be due to the exposure and use of certain data which violates the right to privacy of an individual.

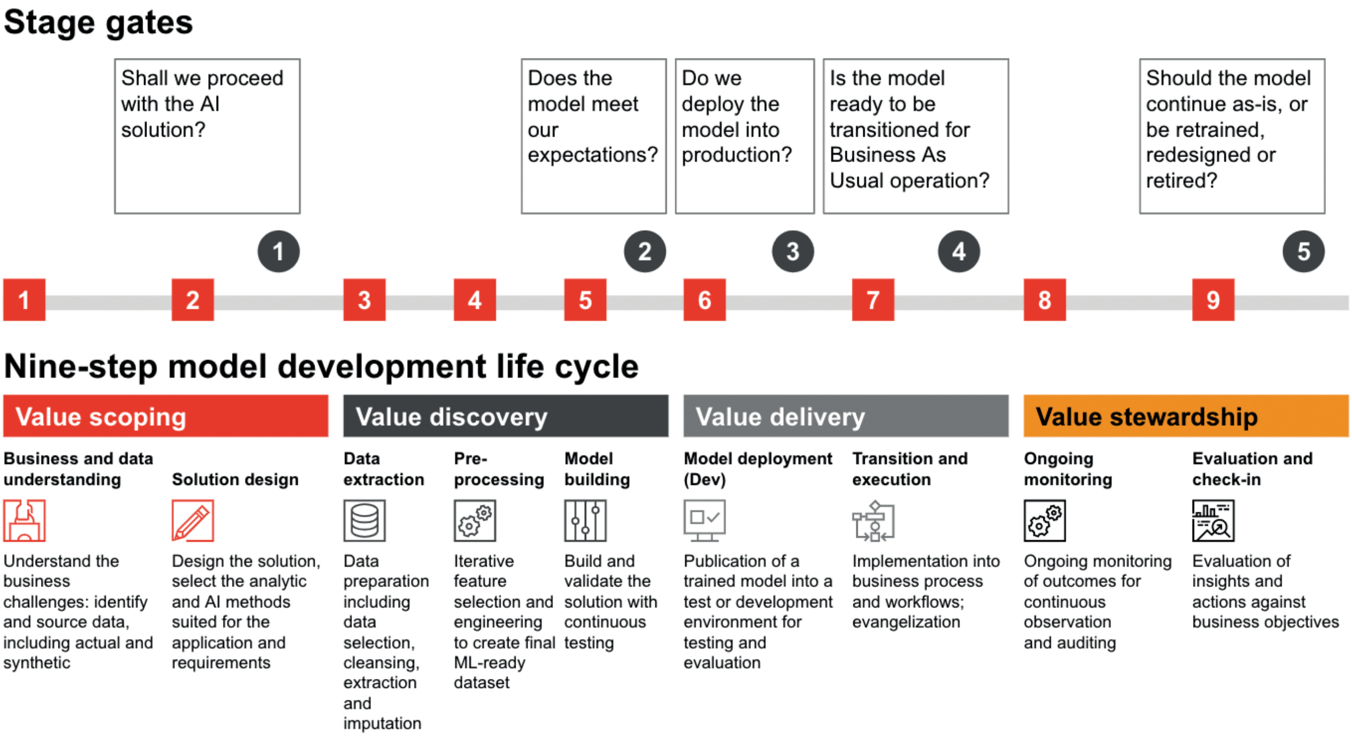
While each of these research initiatives and the risks they seek to address is deserving of entire papers, in this paper we focus on several categories of these technologies that fall into the privacy and security domains, as well as the machine learning development, and how these techniques layer in.

Leveraging AI and analytics to address privacy and security challenges

To understand where privacy and security risks may affect algorithmic systems, we first need to understand the model development life cycle.

The ML development life cycle

A model development life cycle (see Figure 2) comprises the steps between: 1) formulating the problem for a future solution to solve, including success criteria, understanding and accessing data that may be used to train that model; 2) the iterative process of data processing and model training to build a prototype; 3) scaling that prototype to a full-fledged solution; 4) monitoring the solution in production and how users and leveraging the solution; and 5) observing if the model needs to be



Stage gates

Shall we proceed with the AI solution?

Does the model meet our expectations?

Do we deploy the model into production?

Is the model ready to be transitioned for Business As Usual operation?

Should the model continue as-is, or be retrained, redesigned or retired?

Nine-step model development life cycle

Value scoping

Value discovery

Value delivery

Value stewardship

Business and data understanding: Understand the business challenges: identify and source data, including actual and synthetic

Solution design: Design the solution, select the analytic and AI methods suited for the application and requirements

Data extraction: Data preparation including data selection, cleansing, extraction and imputation

Pre-processing: Iterative feature selection and engineering to create final ML-ready dataset

Model building: Build and validate the solution with continuous testing

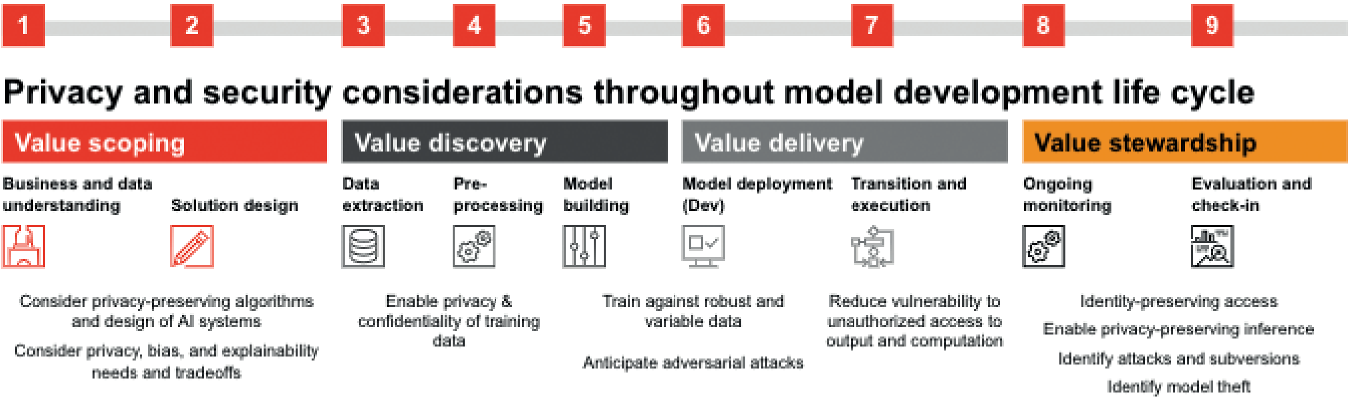
Model deployment (Dev): Publication of a trained model into a test or development environment for testing and evaluation

Transition and execution: Implementation into business process and workflows; evangelization

Ongoing monitoring: Ongoing monitoring of outcomes for continuous observation and auditing

Evaluation and check-in: Evaluation of insights and actions against business objectives

Figure 2: Nine step model development life cycle. Responsible AI: Maturing from theory to practice
Source: PwC[22]



Privacy and security considerations throughout model development life cycle

Value scoping

Value discovery

Value delivery

Value stewardship

Business and data understanding

Solution design

Data extraction

Pre-processing

Model building

Model deployment (Dev)

Transition and execution

Ongoing monitoring

Evaluation and check-in

Consider privacy-preserving algorithms and design of AI systems

Consider privacy, bias, and explainability needs and tradeoffs

Enable privacy & confidentiality of training data

Train against robust and variable data

Anticipate adversarial attacks

Reduce vulnerability to unauthorized access to output and computation

Identity-preserving access

Enable privacy-preserving inference

Identify attacks and subversions

Identify model theft

Figure 3: Privacy and security-enhancing opportunities throughout AI development and use
Source: PwC, 2022

adjusted at all, or even taken offline, once in production.

Throughout this life cycle, there are pain points from a security and privacy point of view.

- What data can we use? How can we make it so that we can use a large enough and meaningful data set that will enable us to build a robust model or process?;
- How do we stress test a model or process against different potential vulnerabilities and sensitivities?;
- How might nefarious actors try to attack my models?;
- How do we leverage existing cyber

security practices for a solution we may only partially own?;
- How are people using this system? Can we detect potential subversions or misuses? Does this solution still work as intended?

To address these questions, several capabilities are emerging to help organisations address these pain points. In the area of privacy, several new technologies are emerging to help developers mitigate privacy risks (see Figure 3). In fact, your technical teams may be asking you about them already. Similarly, new strategies recognising the unique attributes and vulnerabilities of AI are needed to help organisations secure AI systems.

Privacy enhancing technologies (PETs)

PETs are a growing trend due to the convergence of three events: regulatory requirements to protect data and data privacy, the increased need for more granular data to feed large ML and deep learning (DL) systems and demand for increased privacy protections by consumers. The privacy community is increasingly interested in PETs. For example, the UK's data protection regulator, the Information Commissioner's Office (ICO), recently released a whitepaper on the topic.[23] There are several major categories of PETs.

Differential privacy

Differential privacy (DP) is a technique that ensures that anyone using any database for learning will use an approximate version of that database.[24]

Federated learning

Federated learning (FL) is a technique that trains an algorithm across multiple decentralised edge devices or servers using local data samples, without exchanging them.[25]

Homomorphic encryption

Homomorphic encryption (HE) is a technique that permits users to perform computations on its encrypted data without first decrypting it.[26]

Secure multiparty computation and confidential compute

Secure multiparty computation (SMPC) is a technique that splits the data and assigns the data to multiple trusted third parties so that computation can be done on the split data across third parties without sharing data between each other. Related, confidential compute does all the compute on client data servers.[27]

Synthetic data

Synthetic data generation is a data generation and privacy technique for creating statistically similar data that can preserve sensitive data and can be used in ML models when there is a lack of data, or the data is highly sensitive.[28]

These techniques show significant promise, but much more work is to be done before these tools will be widely available for all contexts and uses of data.[29,30] Governance and compliance organisations should remain practical and evaluate the use of these technologies where relevant.

Security for AI and data

By and large, companies' data retention is intended to strengthen their abilities to utilise ML techniques at scale. Security risks for algorithmic systems are in many ways the same as security risks to traditional software (see Figure 4); however, AI systems may have a few additional vulnerabilities.

Securing AI systems and the data that uses them requires knowledge about unique ways in which AI and data can be subverted. A few key themes:

1. *Protecting access to core training data*: Leveraging existing cyber security

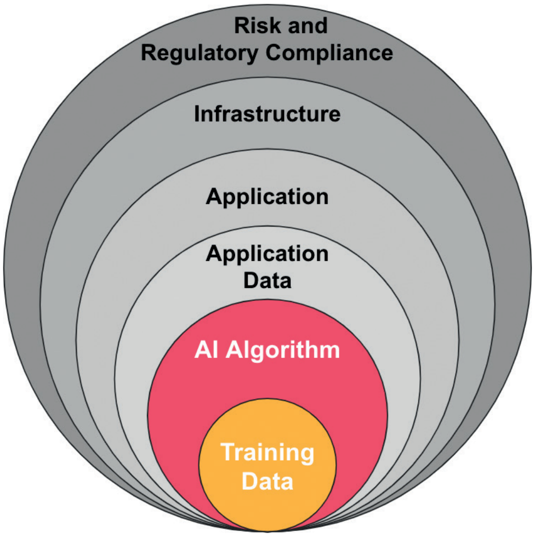


Figure 4: AI product ecosystem
Source: PwC, 2020

practices to decrease access and protect environments;

2. *Data poisoning*: Injecting poor information to a live model, knowing that poor information will be used to retrain a model and worsen its performance;
3. *Model inversion or theft*: Reverse engineering a model based on legitimate queries;
4. *Membership inference*: Unmasking sanitised data or identifying who was included in a data set, which may be a result of combining data sources.

Updated approaches to mitigating security risks need to become part of an emergent governance programme.

Balancing privacy and security risks with other ethics initiatives

Managing risks in organisations is complex, and there is not always a right answer. In mitigating a privacy risk by deleting protected class data, we may inadvertently cause an issue where identifying bias risks becomes more challenging. In all cases, working closely with development teams to mitigate risks also requires a practical approach to balance risks while not instituting overly cumbersome compliance practices. This agility is part of a mature governance and compliance organisation.

BUILDING A HIGH-PERFORMING DATA ETHICS PROGRAMME FROM THE GROUND UP

Ethics and your data protection programme

If you are wondering where to start when building a data and technology ethics programme, consider any existing privacy programme already established at your company. Privacy and data or technology ethics share many similarities — both are contextual and about people. Privacy deals with personal data about an individual, while data and technology ethics deals with algorithms and impacts of AI and data processing that may affect an individual, such as bias or discrimination.

In your privacy programme, you likely have people, processes and tools to manage compliance with GDPR,[31] including its restrictions on automated decision making and requirements to perform privacy impact assessments or data protection impact assessments (DPIAs) and legitimate interest analyses. If you work for a US company, you are probably dealing with similar compliance challenges under CCPA/California Privacy Rights Act (CPRA),[32] Virginia and California laws obligating companies to review automated decision making. Other emerging regulations specific to the use of data in AI/ML systems, like those around biometric privacy in Illinois or a proposed law banning the use of certain data for underwriting purposes in Colorado,[33] continue to evolve, requiring an integrated ethics and governance function.

Integrating ethics

Depending on your organisation structure, legal and business teams managing AI/ML and developing a data and technology ethics programme can seek out partnerships and alignment with compliance teams managing other types of compliance (such as corporate ethics, code of business conduct, bribery) and leverage resources and processes. Alternatively, you may prefer to incorporate data and technology ethics components into your privacy and security programmes. Many companies are starting to take this approach and combine data protection with data ethics under a trust organisation, with a Chief Trust or Chief Compliance Officer as the lead executive of that team or department. Be clear about the purpose of any data ethics programme you intend to launch — it is better to establish an effective right-sized governance framework to scale the responsible and ethical acquisition and use of data for AI/ML projects while maintaining customers' trust and complying with contractual and legal obligations.

Methodology

Take time to discuss the problem statement or data use case and potential solutions with relevant stakeholders and incorporate learnings from any pilot or existing data protection programmes and review committee structures before you start building your data ethics programme. In addition to traditional legal and compliance stakeholders, make sure to include user experience teams to provide insight on UX implications of ML projects.

Identify the current state of your AI/ML data use and programme status. What is the level of maturity? Is it ad hoc, managed, optimised? How is it resourced and funded? Is it integrated or siloed?

Take an inventory of programme components. Do you have data and technology ethics principles, a data and technology ethics policy, playbooks or other training materials, or an intake form for teams to enter details on in-scope projects that can then be reviewed by appropriate legal and programme stakeholders? Do you have a tool to track intake forms and assessments, automate certain tasks and create reports?

What metrics or success criteria do you wish to establish to track and measure goals? A leading practice is to establish a streamlined, compliant and repeatable governance process with identified decision makers and clear escalation path(s) that should include a framework for the use of data, supported by repeatable processes and platform technical capabilities.

Service provider role

There is also an opportunity for companies as AI governance service providers for their customers. How can your company offer value to its customers by offering built-in functionality and compliance solutions for AI/ML and automated processing obligations? For example, a company might provide insight into how an AI service works so that results could be interpreted in a way that helps to prevent bias or offer controls and settings in an AI feature to help enable compliance.

This is ethics by design — incorporating ethical practices in the full life cycle of products and offering

Governance harmonisation

When developing your data ethics programme, consider ways to add right-sized governance by reducing additional 'tax' or administrative burdens on your programme's stakeholders and company employees. A possible solution is to harmonise the different types of governance that have an impact on data processing. This will involve strategic discussions with the leads of your emerging data ethics programme and leads of your privacy and security programmes. You should solicit input from a wide range of backgrounds from legal to compliance, product designers and engineers and data scientists.

First, agree on the target of governance. Is it the project or business initiative or is it preferable to govern at a more granular level such as the data sets in question?

Second, standardise the language used in programme documentation such as policies, standards, guidance and playbooks used for training. Be mindful of the definitions of 'data', 'data set', 'ML, 'personal data' and other subsets of data. Where there are discrepancies, seek to align so there is consistency across your data protection programmes, or call out specific distinctions relevant to one area of the programme.

Third, establish a single technical home (this could be an internal platform or vendor solution) for governance artefacts and process documentation. This repository should be accessible to relevant stakeholders and help drive a sentiment of 'let us all build here' in a compliant, controlled and documented manner. Eventually, create a shared back-end database to enable reporting and auditing functions.

Fourth, include cross-references in any unique intake process for components of your legal review or privacy and security programme review so that they 'speak' to one another. For example, the question 'Are you using personal data?' Answer: 'Yes' Output 'Please complete a PIA' may show up in multiple intake processes. To help reduce overhead and business colleague confusion or avoidance of compliance processes, it is better to eventually conform multiple intake processes and associated documentation so there is one path to compliance and enablement of value-oriented projects.

Finally, create shared evaluation processes and escalation paths. For example, low risk projects that conform to consolidated privacy, security and data ethics guidance and legal review could proceed so long as risks and remediations are documented. Medium to high-risk projects may require additional review by a consolidated cross-functional committee where legal, privacy, security and data ethics issues can be addressed in one process. This reduces the churn of requiring business teams to seek multiple inputs and follow multiple processes and encourages compliance.

Three stages of development

It is useful to envision the development of your data ethics and technology governance programmes in incremental stages. In this way, you can track and keep pace with evolving regulatory requirements and your company's strategic objectives. Progressive stages also help break the overall programme components into manageable chunks; you will likely have several interim governance and processes until more scalable processes and platform technical capabilities can be established on a company-wide level.

Years 1–2: Baby steps/crawl

This is the incubation stage of a programme where you are taking small, deliberate steps to implement a data and technology ethics framework. You might limit this to an initial ethical assessment of an AI/ML data use project. This is the observation and learning stage where you are gathering data about what is needed at your organisation and what can work from an organisational standpoint.

Years 3–5: Starting to walk

In subsequent years, your programme is starting to mature. Your data and technology ethics framework evolves to incorporate ethics by design, training and awareness to drive behaviour aligned with your company culture and values. Development teams are engaged for feedback and to support the development of new practices to scale.

Years 5+: Mature governance — run state

Your programme is all grown up but still needs ongoing maintenance and updates for continuous improvement. Here you might increase internal and external communications about your data and technology ethics programme (for example via blog posts, whitepapers or on a company trust centre) and extend the data and technology ethics framework and ethics by design to *ex-post* review of AI/ML data use projects.

CONCLUSION

Data governance programmes, to be effective, should consider the needs of the teams and business outcomes they are designed to serve, as well as facilitate compliance to emerging regulations and policy. To succeed, these governance programmes should be foundational, and must find a balance between enabling innovation and mitigating risks. Even as the regulatory climate shifts, many organisations see a growing appreciation for AI, technology and data risks above and beyond those specified in existing law; as such, there are thankfully many individuals advancing governance practices and 'ethics by design'. Codified practices can build off these nascent practices for optimal adoption. A champion model may be implemented to allow for the propagation of new requirements as they come to fruition.

Regardless of the approach chosen by a specific organisation, aligning governance practice with technical feasibility is paramount — and technical teams, as such, should be empowered to participate.

ACKNOWLEDGMENT

The authors would like to thank Soojin Jeong and Ambia Harper for their contributions to this paper.

References

1. United States Government Accountability Office (July 2021), 'Facial Recognition Technology', available at https://www.gao.gov/assets/gao-21-105309.pdf (accessed 11th November, 2022).
2. Garvie, C., Bedoya, A. and Frankle, J. (October 2016), 'The Perpetual Line-up: Unregulated Police Face Recognition in America', Georgetown Law Center on Privacy and Technology, available at https://www.perpetuallineup.org/ (accessed 11th November, 2022).
3. Buolamwini, J. and Gebru, T. (2018), 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification', Proceedings of Machine Learning Research, Vol. 81, pp.1–15, available at http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf (accessed 11th November, 2022).
4. Hill, K. (December 2020), 'Another Arrest, and Jail Time, Due to a Bad Facial Recognition Match', *New York Times*, available at https://www.nytimes.com/2020/12/29/technology/facial-recognition-misidentify-jail.html (accessed 11th November, 2022).
5. Ahmed, S., Nutt, C. T., Eneanya, N. D., Eneanya, N. D., Reese, P. P., Sivashankar, K., Morse, M., Sequist, T. and Mendu, M. L. (2021), 'Examining the Potential Impact of Race Multiplier Utilization in Estimated Glomerular Filtration Rate Calculation on African-American Care Outcomes', *Journal of General Internal Medicine*, Vol. 36, pp. 464–471.
6. Hughes, D. (August 2020), 'What is the A-level algorithm? How the Ofqual's grade calculation worked – and its effect on 2020 results explained', i, available at https://inews.co.uk/news/education/a-level-algorithm-what-ofqual-grades-how-work-results-2020-explained-581250 (accessed 11th November, 2022).
7. United States Government Accountability Office, ref. 1 above.
8. Mayson, S. (2019), 'Bias in, Bias Out', *The Yale Law Journal*, available at https://www.yalelawjournal.org/pdf/Mayson_p5g2tz2m.pdf (accessed 11th November, 2022).
9. Obermeyer, Z., Powers, B., Vogeli, C. and Mullainathan, S. (October 2019), 'Dissecting racial bias in an algorithm used to manage the health of populations', *Science*, Vol. 366, No. 6464, pp. 447–453.
10. Intersoft Consulting, 'General Data Protection Regulation, Art. 22 Automated individual decision-making, including profiling', available at https://gdpr-info.eu/art-22-gdpr/ (last accessed 11th November 2022)
11. Intersoft Consulting, 'General Data Protection Regulation, Art. 25 Data protection by design and by default', available at https://gdpr-info.eu/art-25-gdpr/ (accessed 11th November, 2022).
12. Intersoft Consulting, 'General Data Protection Regulation, Art. 12 Transparent information, communication and modalities for the exercise of the rights of the data subject', available at https://gdpr-info.eu/art-12-gdpr/ (accessed 11th November, 2022).
13. Intersoft Consulting, 'General Data Protection Regulation, Art. 35 Data protection impact

assessment', available at https://gdpr-info.eu/art-35-gdpr/ (accessed 11th November, 2022).

14. European Commission (2021), 'Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS, 21st April 2021', available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206 (accessed 11th November, 2022).
15. California Legislative Information, 'TITLE 1.81.5. California Consumer Privacy Act of 2018 [1798.100 – 1798.199.100] (Title 1.81.5 added by Stats. 2018, Ch. 55, Sec. 3)', available at https://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?division=3.&part=4.&lawCode=CIV&title=1.81.5 (accessed 11th November, 2022).
16. Jillson, E. (April 2021), 'Aiming for truth, fairness, and equity in your company's use of AI', Federal Trade Commission, available at https://www.ftc.gov/business-guidance/blog/2021/04/aiming-truth-fairness-equity-your-companys-use-ai (accessed 11th November, 2022).
17. 117th Congress (2021/22), 'H.R.2154 – Protecting Americans from Dangerous Algorithms Act', available at https://www.congress.gov/bill/117th-congress/house-bill/2154/text (accessed 11th November, 2022).
18. 117th Congress (2021/22), 'H.R.5596 – Justice Against Malicious Algorithms Act of 2021', available at https://www.congress.gov/bill/117th-congress/house-bill/5596/text (accessed 11th November, 2022).
19. 117th Congress, 2nd Session (July 2022), 'H.R.8152', available at https://energycommerce.house.gov/sites/democrats.energycommerce.house.gov/files/documents/BILLS-117hr8152ih.pdf (accessed 11th November, 2022).
20. The Institute of Internal Auditors (IIA) (2020), 'The IIA's Three Lines Model', available at https://www.theiia.org/globalassets/site/about-us/advocacy/three-lines-model-updated.pdf (accessed 11th November, 2022).
21. EAIDB, 'The Ethical AI Database', available at https://www.eaidb.org (accessed 11th November, 2022).
22. PwC (2021), 'Responsible AI – Maturing from Theory to Practice', available at https://www.pwc.com/gx/en/issues/data-and-analytics/artificial-intelligence/what-is-responsible-ai/pwc-responsible-ai-maturing-from-theory-to-practice.pdf (accessed 11th November, 2022).
23. ICO (September 2022), 'ICO publishes guidance on privacy enhancing technologies', available at https://ico.org.uk/about-the-ico/media-centre/news-and-blogs/2022/09/ico-publishes-guidance-on-privacy-enhancing-technologies/ (accessed 11th November, 2022).
24. Cheruvu, R. (November 2018), 'A High-Level Introduction to Differential Privacy', Towards Data Science, available at https://towardsdatascience.com/a-high-level-introduction-to-differential-privacy-edd20e6adc3b (accessed 11th November, 2022).
25. Kelvin (October 2020), 'Introduction to Federated Learning and Challenges', Towards Data Science, available at https://towardsdatascience.com/introduction-to-federated-learning-and-challenges-ea7e02f260ca (accessed 11th November, 2022).
26. Thaine, P. (December 2018), 'Homomorphic Encryption for Beginners: A Practical Guide (Part 1)', Medium, available at https://medium.com/privacy-preserving-natural-language-processing/homomorphic-encryption-for-beginners-a-practical-guide-part-1-b8f26d03a98a (accessed 11th November, 2022).
27. Inpher, 'What is Secure Multiparty Computation?', available at https://inpher.io/technology/what-is-secure-multiparty-computation/ (accessed 11th November, 2022).
28. Tan, L. (February 2021), 'Synthetic Data: Applications in Data Privacy and Machine Learning', Towards Data Science, available at https://towardsdatascience.com/synthetic-data-applications-in-data-privacy-and-machine-learning-1078bb5dc1a7 (accessed 11th November, 2022).
29. Privacy Tech Alliance/Future of Privacy Forum (June 2021), 'Privacy Tech's Third Generation', available at https://fpf.org/wp-content/uploads/2021/06/FPF-PTA-Report_Digital.pdf (accessed 11th November, 2022).
30. Garrido, G. M., Sedlmeir, J., Uludag, Ö., Alaoui, I. S., Luckow, A. and Matthes, F. (July 2022), 'Revealing the Landscape of Privacy-Enhancing Technologies in the Context of Data Markets for the IoT: A Systematic Literature Review', *Arxiv*, 2107.11905.
31. European Union, 'Complete guide to GDPR compliance', available at https://gdpr.eu (accessed 11th November, 2022).
32. State of California Department of Justice, 'California Consumer Privacy Act (CCPA)', available at https://oag.ca.gov/privacy/ccpa (accessed 11th November, 2022).
33. Colorado General Assembly (2021), 'Restrict Insurer's Use of External Consumer Data', available at https://leg.colorado.gov/bills/sb21-169 (accessed 11th November, 2022).