FEATURE

PRIVACY CHALLENGES IN A WORLD OF BIG DATA, ML AND AI

BY **FRASER TENNANT**

Technology innovations such as Big Data, artificial intelligence (AI) and machine learning (ML) are among the most advanced to have emerged in recent years, bringing substantial benefits for society and consumers. However, with great advancement comes great responsibility, and privacy issues are among the most challenging that organisations face.

"Data is used in every facet of business," says Martin Owen, vice president of Erwin. "Big Data provides a vehicle for organisations to use vast amounts of data without necessarily structuring its use, and AI and ML look to make sense of it by providing algorithms and context. That said, privacy is critical, as technology allows data to be used anywhere, by anybody."

Indeed, a 2017 Information Commissioner's Office (ICO) report – 'Big data, artificial intelligence, machine learning and data protection' – characterises privacy as an "enabling right", a right that enables not only societal benefits such as dignity, personality and community, but also organisational benefits like creativity, innovation and trust. These benefits, according to the report, will be strengthened under the new General Data Protection Regulation (GDPR), which requires organisations to be more transparent and accountable for what they do with personal data – a requirement that applies to Big Data, AI and machine learning.

"Privacy is a major issue across the globe at the moment and rightly so," says Chris Gayner, director of labs at Symphony Ventures. "From a citizen's perspective, the implications include data theft, financial fraud and loss of data. From a business perspective, we can add reputation damage – particularly for publicly traded companies in which a major impact in brand or commercial reputation can have a detrimental impact on share price or trading relations."

In the view of Dr Alexander Stein, founder of Dolus Advisors, it is the combination of privacy and technology that is the hub of the matter – two disparate elements which only entwine as a result of the technological age in which we live. "But the core challenges – tensions between individual, institutional and governmental understandings of privacy – remain fundamentally as they have been since the dawn of civilization," he says. "One discrete problem is the confusion and lack of consensus about how to regulate and manage two different things which have now become radically combined faster and in ways businesses and society are inadequately prepared for."

The complexity of privacy

Drilling down further into the privacy implications of Big Data, it is clear that it is the complexity of Big Data analytics as well as AI and ML techniques that are serving to make things difficult for organisations.

"In addition to the privacy issues raised by Big Data, AI presents new problems," suggests Omer Tene, vice president and chief knowledge officer at the International Association of Privacy Professionals (IAPP). "First, algorithmic opacity: as algorithms become complex and ML more iterative, it is increasingly difficult to explain decision making. Second, risks of discrimination: automated decisions may reflect and deepen existing societal bias. Third, lack of due process: individuals may have little insight into or redress against machine-made decisions."

David Tomblin, business analytics engagement specialist at MHR Analytics, sees the main challenge for organisations being whether data handling is fair. "Fairness involves several elements but the key is transparency – being able to prove what, where and why personal information is being held and processed," he says. "In theory, this could cause challenges in Big Data systems as they typically involve repurposing data in unpredicted ways, using multifaceted algorithms to create scenarios about individuals, with unexpected and sometimes, especially in the context of privacy, unwelcome effects."

According to Parnian Najafi, senior cyber security analyst at FireEye, even if organisations adopt strict privacy policies, the likelihood is that cyber hackers will ignore them and continue to capture data via alternative, advanced technologies. "We cannot have privacy without security," she asserts. "Although organisations try to protect the privacy and security

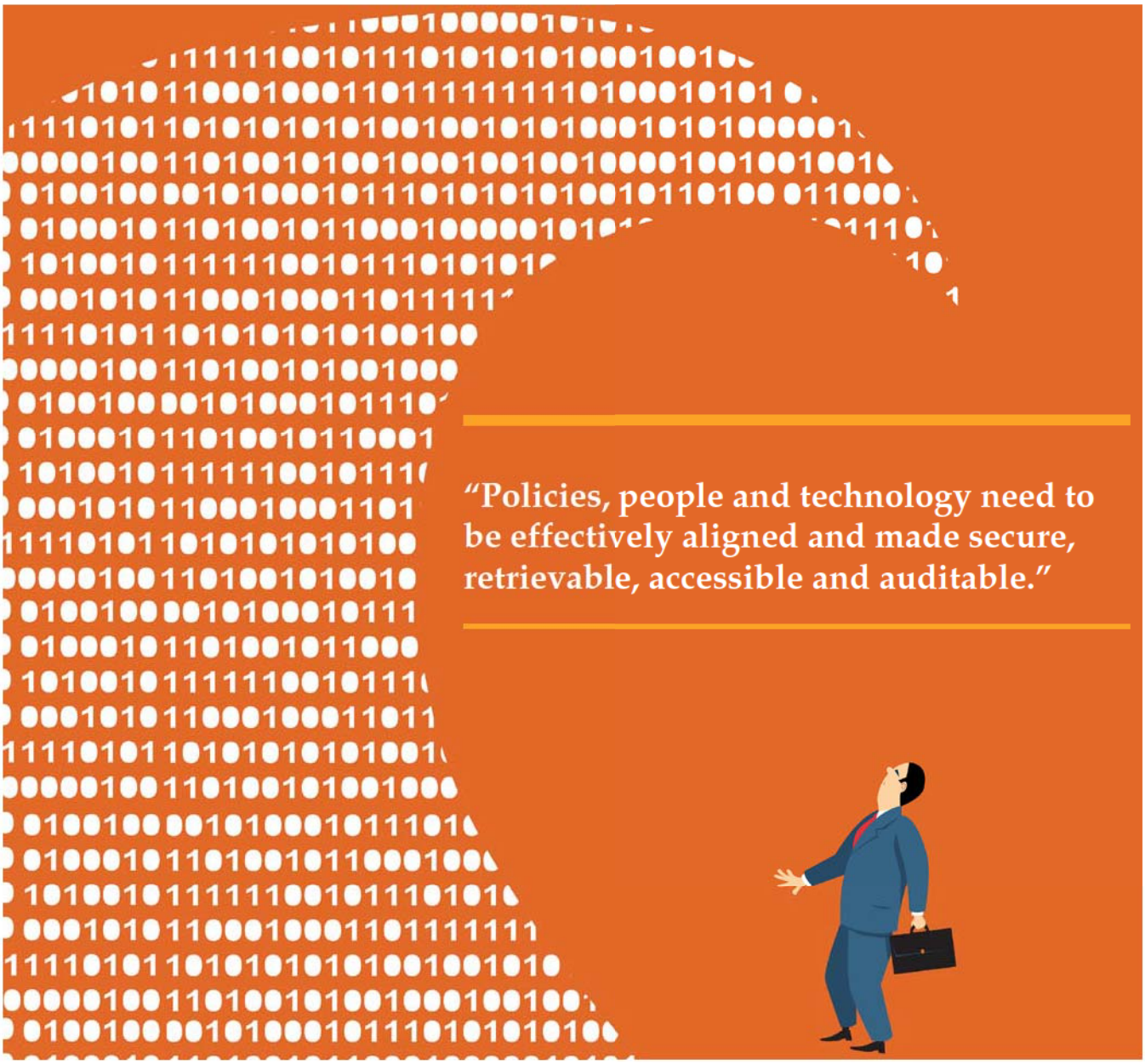
of user data, the amount of aggregated data and the information extracted through AI algorithms could make data breaches even more damaging. Furthermore, even if users' private data is used for training data in ML models, they can be reverse engineered, jeopardising users' privacy."

Another significant problem, according to Dr Stein, is that the psychological dimensions of privacy and security are being subordinated under procedural, technological, legal, commercial and policy considerations. "The unintended adverse consequences of this cannot be understated," he suggests. "There is a discernible cause-and-effect correlation between increasing losses of control over privacy and personal data and the escalating volume and scale of data breaches and other malicious incidents."

Encryption and containerisation

While the privacy challenges posed by Big Data, AI and ML are considerable, the good news is that there are a number of tools available, such as encryption and containerisation, which can assist organisations. To this end, policies, people and technology need to be effectively aligned and made secure, retrievable, accessible and auditable.

"The level of encryption, of course, varies depending on the type of data and thus the point at which encryption occurs needs to be factored into a wider data management strategy," advises Mr Gayner. "As technologies such as automation and AI become



"Policies, people and technology need to be effectively aligned and made secure, retrievable, accessible and auditable."

more prevalent, a robust data strategy that takes into account on-premise, cloud and hybrid infrastructures will be critical."

While agreeing that encryption and containerisation tools are important when seeking to implement data privacy and security, Mr Owen points out that, without the infrastructure and process design of systems and data, such tools do not have the context of what to encrypt or to put into containers. "Encryption and containerisation are not the end goals, but are tools to achieve privacy and security," he says. "Other tools such as security event and incident management (SEIM) are equally important in visualising and monitoring the state of our security."

As far as additional privacy protection tools are concerned, Mr Tene notes that computer scientists are developing new methodologies, such as differential privacy, a formal definition of privacy which allows for rigorous tradeoffs against utility, secure multiparty computations and homomorphic encryption, which allow organisations to conduct data analysis without revealing secret information, and blockchain.

That said, according to Dr Stein, "there are no universal tools or systems for ensuring data privacy and security, as different organisations capture and use data differently. These differences also mean that organisations are uniquely vulnerable. Ultimately, information is a commodity, however differently it is valued by any given business. Secure data is primarily the product of organisational integrity, principled leadership and ethical governance. Organisations need to adopt and internalise privacy as a key institutional imperative, equal in importance to shareholder value, brand reputation and products and services standards. Privacy by design must have its place alongside security by design."

Privacy and the GDPR

The GDPR, which came into force on 25 May 2018, is expected to revolutionise the way data is processed across all platforms, especially where privacy is concerned. And while the GDPR is a European Union (EU) regulation, it encompasses entities that do business in the EU or market to EU citizens, even if said entities are located in countries outside the territory.

"The GDPR can help companies develop privacy guidelines for their employees and customers," suggests Ms Najafi. "It is also forcing global companies to have a company-wide privacy compliance programme to avoid fines associated with non-compliance. While some companies may not have focused on privacy issues, these regulations may force them to start or mature their existing privacy compliance programme. They may also hire new employees to specifically focus on privacy compliance."

For Mr Owen, the GDPR will compel organisations to get a grip on customer and employee data and, although the legislation is not yet prescriptive, force them to view information security as critical.

"Regulations such as the GDPR will transform the culture of the way we do business," he believes. Big Data, AI and business intelligence will not be actionable without some form of data or information governance approach from a company. Enterprises need to understand that their data is trustworthy, from a trusted source, is secure and is what they think it is."

In Mr Gayner's experience, forward-thinking companies are looking toward automation and AI tools to enhance the process elements of GDPR, such as data subject access rights (DSAR). "These technologies bring additional benefits of speed, reliability, quality and auditability," he says. "However, the majority of executives are simply not thinking this way and are instead building an army of administrators to handle the process challenges. This is an unsustainable solution in light of ever-evolving regulation."

Evolving technologies

With a raft of new technologies likely to emerge in the years ahead, the privacy issues facing organisations seem set to increase exponentially. Many of these will be as a result of the rapidly growing use of AI and concurrent uptick in personally identifiable information (PII).

"As technology becomes progressively complicated, it is necessary for regulations, security and privacy solutions to catch up," observes Ms Najafi. "Organisations should ensure that they comply with privacy laws. Transparency is important when discussing how the data will be used, how long they plan to store such data and whether they will share it with other companies. Overall, organisations have to consider the broader implication of their decisions."

While some organisations do inform users about data collection and sharing practices in their privacy policies, they are often difficult to comprehend. In this regard, AI-based models – which provide an automatic and comprehensive framework for privacy policies analysis – can help users to better understand the scope of such policies.

The evolving technologies that are most likely to impact the data privacy space in the years ahead will be biometrics and the maturity of the IoT, according to Mr Tomblin. "The biometrics arena is on the march, with fingerprints and now facial recognition becoming commonplace on smartphones," he explains. "Look out for usage extending into the smart card industry with the offer of an alternative to PIN security. The next generation of vehicles and autonomous driving will also be one to watch. If when travelling we have our hands free, this will only lead to more ways to increase our digital profiles and feed more personal data into Big Data systems."

Loosening straightjackets

Keeping track of when, why, where and how data is collected, used, stored and reported is clearly a major challenge. Moreover, with many organisations

lacking a cohesive data privacy strategy – a scenario fostered largely by having little sense of urgency or even awareness of the threats they face – operations are in danger of being compromised.

"Looking ahead, the resolution to our legitimate concerns regarding the preservation and protection of personal privacy will not be purely technical or mechanistic," says Dr Stein. "No matter the technological advancements thrown at these problems, however innovative and dazzling, nothing will improve until we consider the intended and unintended consequences of our decisions and actions."

In a world in which the speed of technological innovation shows little sign of abating, the overarching challenge for organisations across the globe is most assuredly this: to establish a data security regime that avoids straightjacketing operations and ensures that privacy does not become a casualty of technology. R&C