

Privacy Engineering

Joseph Williams^{ID} and Lisa Nee, Infosys Consulting

Privacy has moved beyond the “by design” phase to how we actually need to engineer the solutions.

Much of the current thinking regarding the roles of the Chief Privacy Officer (CPO) and the Data Protection Officer (DPO) evolved from the last decade’s rollout of the European Union’s (EU’s) General Data Protection Regulation (GDPR). Early GDPR concerns initially focused on “big picture skills” and administrative expertise¹ that prioritized legal and policy analyses of GDPR (and, subsequently, additional regulations such as the California Consumer Privacy Act). A half-dozen years later, frameworks have evolved for implementing privacy policies, procedures, and governance, the most common including some variation of the Privacy by Design (PbD) philosophy approach promulgated by Ann Cavoukian.² Now there is a significant upshift in focus to what needs to be specifically implemented to achieve the outcomes promised by those frameworks. This upshift has given new traction to the specialty of privacy engineering, which is the technical companion to the policy roles played by the CPO and the DPO.

The shortcomings in privacy engineering are expensive. Businesses operating in Europe were fined US\$1.2 billion in 2021 for violations of GDPR privacy regulations.³ Research shows that business suffering data breaches that impact private data can result in a permanent loss of 21% of their customers.⁴

Privacy is now a complicated ecosystem for all enterprises. The diverse rubric of state, national, and regional legal requirements make compliance a challenge. Many customers have an expectation of privacy that is an important factor in their willingness to establish and maintain business relationships. Privacy architectures need to be flexible and yet robust—no easy feat.

Privacy engineering is an emerging field that develops the tools, methodologies, and processes for meeting the privacy requirements and expectations of regulators and customers. Privacy is intertwined with technologies and techniques for data protection and cybersecurity, so it is much more than a legal or policy issue—it is a technical problem requiring the application of sophisticated solutions that meet legal and customer success objectives.⁵

Translating legal and policy mandates to an appropriate application of technology solutions requires privacy engineering to support the objectives of the Chief Privacy

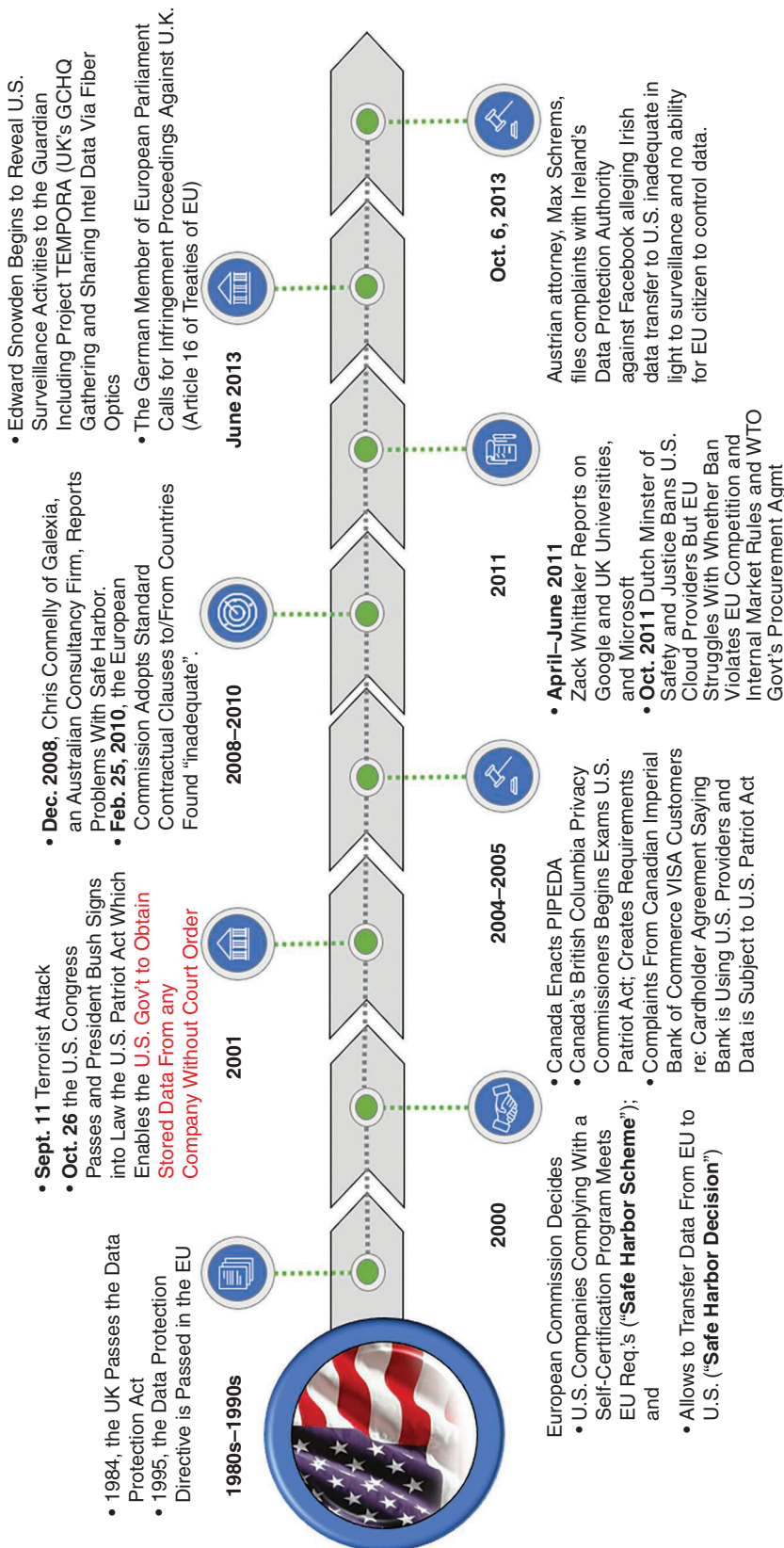


FIGURE 1. A timeline view of the events influencing the development of privacy regulations.

Officer.⁶ At a global enterprise scale, the application of artificial intelligence (AI) technologies to privacy engineering is necessary to effectively protect the enterprise while continuing to enable the productive use of data.

This article provides a survey on why privacy engineering is important, what it means, and some examples of how it can be applied in various architectural constructs. In addition, the article closes with an observation of future risks and opportunities for privacy engineering that needs attention and planning now.

BACKGROUND

The concept of privacy is not new, and its conceptual roots go back to at least Plato and Aristotle. The idea of a right to privacy is more recent and finds its roots in the 19th century when the proliferation of newspapers and sensationalism led to the construction of privacy laws and the inference of privacy rights that found their expression in the famous 1890 *Harvard Law Review* "right to privacy" article by Samuel Warren and Louis Brandeis.⁷ The idea of an individual privacy right is not unique to Western Europe and the United States, but it is there where it has the most history and tradition.

The original concept of privacy concerned itself with protecting one's home and hearth identify from outsiders. In the modern era, privacy concerns have rapidly evolved to include one's digital identity. It is not surprising, given its historical interests in individual privacy, that Western Europe and the United States are leading the charge in protecting digital privacy rights. The recent rush to protect individual privacy rights stems from a series of policy recommendations and regulations that emerged from the rise of large-scale data

collection, transborder data flows, and cybersecurity leaks; collectively, these events led to extensive legislation. Figure 1 provides a timeline of the early events that produced the current GDPR regulatory environment, starting in 2016.

The pronouncements about and regulations of digital privacy rights have produced a set of principles and solution designs for privacy, as depicted in Figure 2. A detailed examination of Privacy Concepts⁸ and PbD² have been addressed elsewhere and represent best practices. The regulatory and reputational consequences for privacy failures are significant, as exemplified in Figure 2. The operational elements of privacy are not well understood, however, which has contributed to the rise of privacy engineering as a way to systematically apply technology solutions to privacy challenges.

As soon as the EU's GDPR was adopted in 2016, experts started worrying about how to comply and how to provide evidentiary support for that compliance.⁹ Privacy engineering, informed by a comprehensive understanding of privacy concepts and application of privacy design principles, provides a rational framework for compliance. Moreover, privacy

engineering, thoughtfully deployed, enables a flexible, responsive, and extensible approach to meeting future regulatory or customer requirements.

The failure to take an engineering approach to privacy makes it virtually

privacy risks. Privacy engineering is based on an approach to systems engineering that leverages a framework such as PbD and tools such as a privacy impact assessment (PIA) throughout the development lifecycle of a program

Privacy engineering is an emerging field that develops the tools, methodologies, and processes for meeting the privacy requirements and expectations of regulators and customers.

impossible to show a strategically thoughtful approach to regulatory compliance. A piecemeal set of point solutions might not satisfy any requirements for responsible intent as required under GDPR Article 25.¹⁰ Privacy engineering also creates value for the enterprise as it can productively influence product design, improve customer trust, and satisfy requirements demanded by other players in the supply chain.¹¹

WHAT IS PRIVACY ENGINEERING?

In its simplest form, privacy engineering is a discipline that takes into consideration privacy principles when creating technical solutions that will mitigate

or system. PbD is typically distilled to seven principles (Figure 3). PbD guides privacy engineers, who create then apply methodologies and tools that provide the requirements architecture. A PIA is a common tool used for identifying and assessing privacy risks throughout the develop lifecycle of a program or system. In general, it analyzes how personally identifiable information is collected, used, shared, and maintained, with the purpose of demonstrating conscious incorporation of privacy protections throughout the develop lifecycle.

Privacy engineers have a variety of techniques they can use to mitigate privacy risks. Each technique has

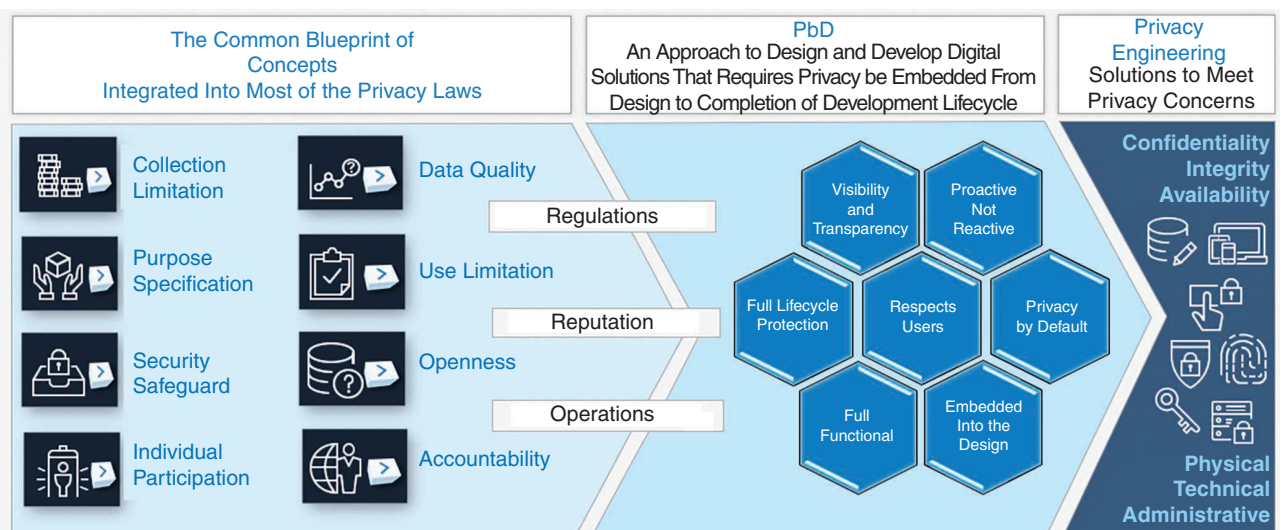


FIGURE 2. The privacy principles that underpin privacy engineering.

a suite of enabling technologies that can be implemented to produce desired outcomes (Figure 4). None of the techniques are perfect, but applied appropriately, they should provide protection for the enterprise and for the privacy of the consumer.

SOME ARCHITECTURAL APPROACHES TO PRIVACY ENGINEERING

There is inherent tension inside the enterprise among the various stakeholders who are involved with data. The CPO, likely an attorney, wants to protect the

enterprise by ensuring data handling meets regulatory requirements, and thus sees data as a liability. Some functions, such as payroll, have statutory requirements to retain data and may see data through the lens of records management. Teams in sales and marketing view data as providing opportunities for additional monetization. Product engineers are looking to data to provide insights into how to improve their offerings. Cross-functional teams want the data to help feed their machine learning models. DPOs are worried about how to manage and control the data. Partnership teams may want to share the data.

Privacy engineering must take into account all their interests and obligations. It makes the architecture of privacy frameworks an interesting challenge.

One interesting approach to privacy engineering that meets many of the aforementioned interests involves anonymizing the data and using them to generate synthetic data that protect the privacy interests of consumers while concurrently enabling the production

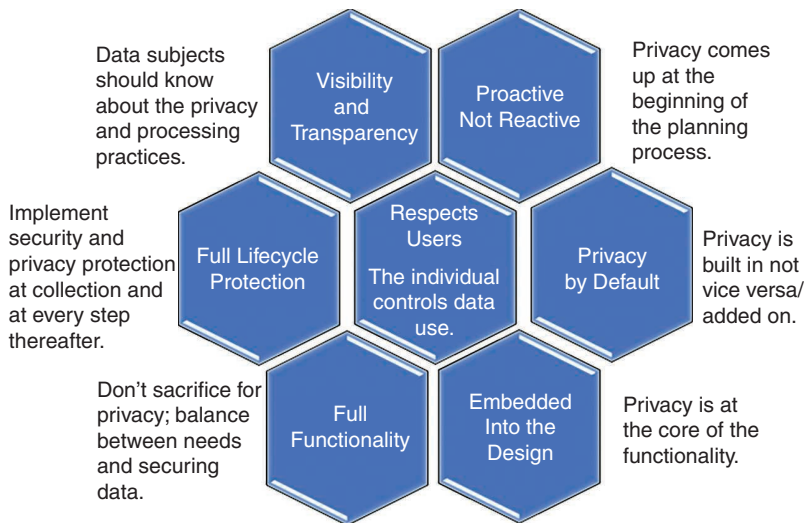


FIGURE 3. The seven PbD principles.



Masking/Deidentifying 	<p>Pseudonymization/tokenization: Replacing a real with a made-up ("pseudo") value or a real value with a made-up value.</p> <p>Hashing/Scrambling Encryption: data are algorithmically scrambled, and only those with access to the appropriate key can view the encrypted data.</p> <p>Noise/Differential Privacy: statistical technique that introduces errors by randomly misclassifying values of categorical variable(s).</p> <p>Perturbation: replacing sensitive info with realistic but inauthentic data, or modifying original data based on predetermined masking.</p> <p>Swapping/shuffling: data for variables from another record so that the data user doesn't know whether the real data values correspond to certain records.</p>	<p>PROS</p> <p>Usually retains functional usability of the data while concealing/ minimizing reidentification risks</p> <p>CONS</p> <ul style="list-style-type: none"> • Sometimes decrease accuracy/ validity of data • Not usable for small data sets • Reverse engineering risks exists for masking algorithms
Blurring 	<p>Aggregation: combining individual subject data with a sufficient number of other subject to disguise the attributes of a single subject (for example, reporting a group average instead of an individual value).</p> <p>Generalization: collecting or reporting values in a given range (for example, using age or age range instead of date of birth), including individual data as members of a set (for example, creating categories that incorporate unique cases) or reporting rounded not exact values.</p> <p>Pixelation: modifying or obscuring visual information (for example, blurring faces in a photograph).</p>	<p>PROS</p> <p>Minimizes risk of identification by focusing on collective, rather than individual data</p> <p>CONS</p> <ul style="list-style-type: none"> • Generally, not useful with a small pool of subjects • Decreases reliability of data; potential for false conclusions

FIGURE 4. Data techniques organized by what each "does."

of useful insights. The narrative and techniques for synthetic data have been dealt with elsewhere¹² and are summarized in Figure 5. Although there are still debates regarding how effectively synthetic data can be used, it seems clear they will have application in the training of AI models.

In addition to having a bag of techniques at their disposal, privacy engineers are developing new architectures to enable more security privacy while improving the ability to use or share that data. Not all data are sensitive or need protection. An increasingly attractive approach to privacy engineering involves the use of a data privacy vault construct that isolates and protects data. When combined with data-anonymization techniques, it also has the potential to make the data more useful. There are also engineering benefits to a privacy vault approach¹³ as it enables targeted protection instead of having to secure the entire data chain, as depicted in Figure 6. A data privacy

vault also provides a centralized and extensible approach to security and regulatory compliance.

The privacy engineering paradigm that has gotten the most traction lately involves leveraging an enterprise's

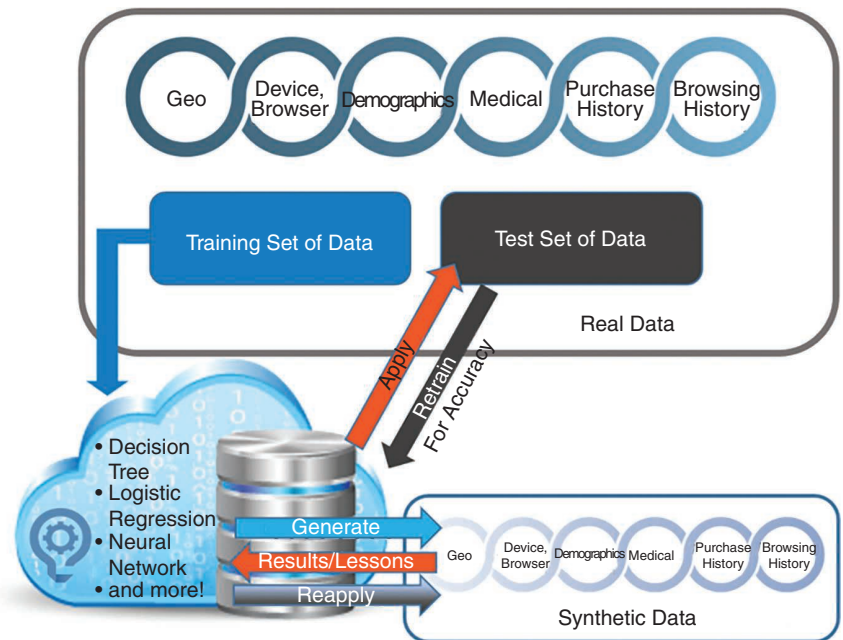


FIGURE 5. A synthetic data overview.

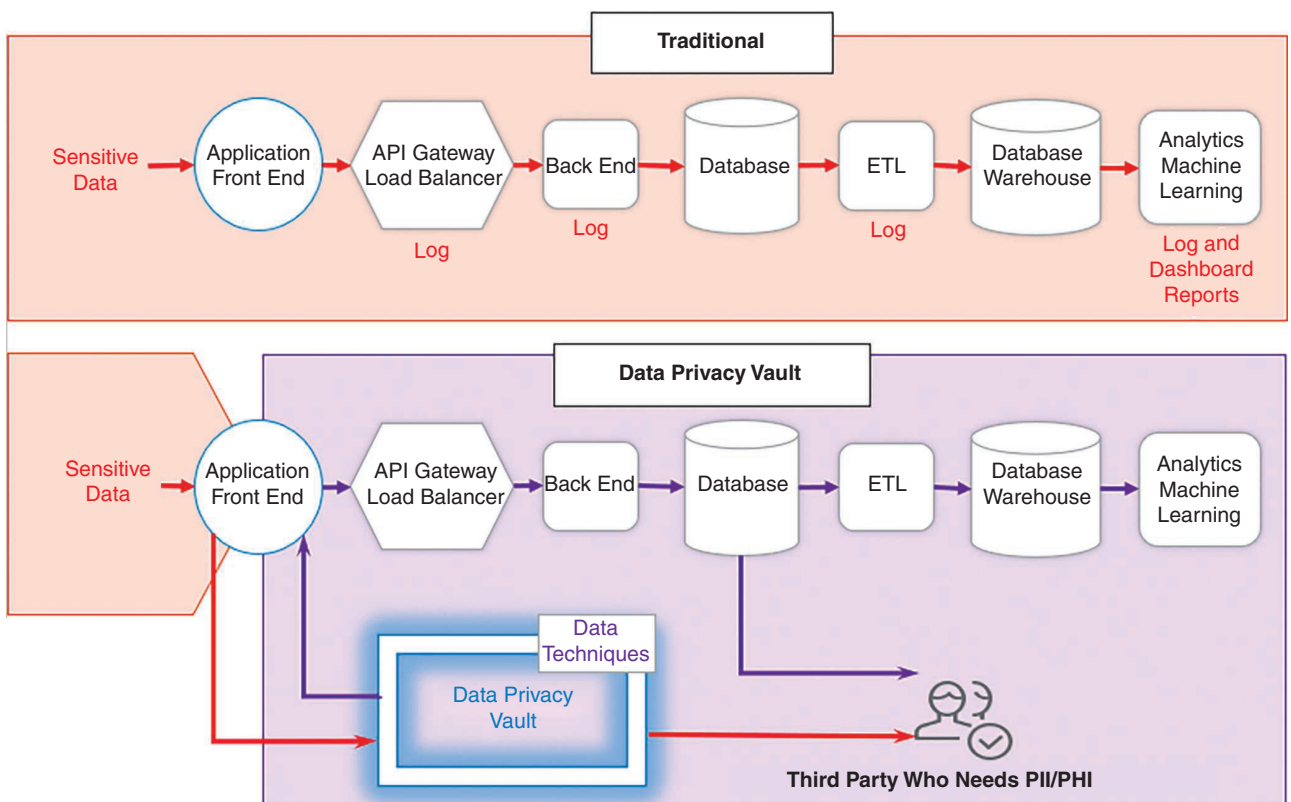



FIGURE 6. Traditional data security versus a data privacy vault. API: application programming interface; ETL: extract, transform, load.

zero-trust architecture (ZTA). Researchers at the National Institute of Standards and Technology have promulgated a collection of concepts (known as SP 800-207) that reduce uncertainty when enforcing access decisions that impact cybersecurity.¹⁴ Privacy engineers have increasingly embraced a ZTA approach, which actually complements a data privacy vault architecture and leverages the value created by the use of synthetic data.

A core technology used in existing privacy engineering frameworks is encryption, which secures data to prevent unauthorized access or surveillance. Quantum computing will eventually provide a means to break the current public-key infrastructure, which is a backbone of encryption. This threat vector has given rise to a number of new cryptographic algorithms that resist quantum computer attacks. Updating of the current asymmetric encryption and signing algorithms, such as those offered by RSA and ECC, need to begin very soon if privacy systems are going to be ready for the advent of quantum computing. 

REFERENCES

1. S. McDougall. "The role of the chief privacy officer in 2020." CPO Magazine. Accessed: Jun. 11, 2022. [Online]. Available: <https://www.cpomagazine.com/data-privacy/role-chief-privacy-officer-2020/>
2. A. Cavouikian, "Privacy by design: The 7 foundational principles," International Association of Privacy Professionals, Portsmouth, NH, USA, Jan. 2011. Accessed: Jul. 16, 2022. [Online.] Available: <https://iapp.org/resources/article/privacy-by-design-the-7-foundational-principles/>
3. M. Schwartz. "Privacy fines: GDPR sanctions in 2021 exceeded \$1 Billion," BankInfoSecurity. Accessed: Jun. 11, 2022. [Online]. Available: <https://www.bankinfosecurity.com/privacy-fines-gdpr-sanctions-in-2021-exceeded-1-billion-a-18331>
4. F. Truta. "Businesses can lose half of customers after a data breach, research shows." Bitdefender. Accessed: Jun. 13, 2022. [Online]. Available: <https://businessinsights.bitdefender.com/businesses-can-lose-up-to-58-of-customers-after-a-data-breach-research-shows>
5. J. Coseglia. "Coffee with privacy pros: DPO vs. CPO, lawyer vs. technician. The dualities of privacy." CPO Magazine. Accessed: Jun. 11, 2022. [Online]. Available: <https://www.cpomagazine.com/data-privacy/coffee-with-privacy-pros-dpo-vs-cpo-lawyer-vs-technician-the-dualities-of-privacy/>
6. S. Gurses and J. M. del Alamo, "Privacy engineering: Shaping an emerging field of research and practice," *IEEE Security Privacy*, vol. 14, no. 2, pp. 40–46, Mar. 2016, doi: 10.1109/MSP.2016.37.
7. S. Warren and L. Brandeis, "The right to privacy," *Harvard Law Rev.*, vol. 4, no. 5, pp. 193–220, Dec. 1890, doi: 10.2307/1321160.
8. "Global comprehensive privacy law mapping chart," International Association of Privacy Professionals, Portsmouth, NH, USA, 2022. Accessed: Apr. 2022. [Online]. Available: <https://iapp.org/resources/article/global-comprehensive-privacy-law-mapping-chart/>
9. E. Edwards. "New rules on data protection pose compliance issues for firms." *The Irish Times*. Accessed: Jul. 12, 2022. [Online]. Available: <https://www.irishtimes.com/business/technology/new-rules-on-data-protection-pose-compliance-issues-for-firms-1.3397742>
10. C. Fennessy, "Privacy engineering: The what, why and how," International Association of Privacy Professionals, Portsmouth, NH, USA, Aug. 8, 2019. [Online]. Available: <https://iapp.org/news/a/privacy-engineering-the-what-why-and-how/>
11. N. Coca. "How data privacy regulations affect your supply chain." Triple Pundit. Accessed: Jun. 14, 2022. [Online]. Available: <https://www.triplepundit.com/story/2018/how-data-privacy-regulations-affect-your-supply-chain/11386>
12. E. Strickland. "Are you still using real data to train your AI?" *IEEE Spectrum*. Accessed: Jul. 6, 2022. [Online]. Available: <https://spectrum.ieee.org/synthetic-data-ai>
13. S. Falconer. "Why everyone needs a data privacy vault." *Software Engineering Daily*. Accessed: Jul. 12, 2022. [Online]. Available: <https://softwareengineeringdaily.com/2022/03/12/why-everyone-needs-a-data-privacy-vault/>
14. S. Rose, O. Borchert, S. Mitchell, and S. Connelly, "Zero trust architecture," National Institute of Standards and Technology, Gaithersburg, MD, USA, NIST Special Publication 800-207, Aug. 2020.

JOSEPH WILLIAMS is partner, cybersecurity practice, with Infosys, Seattle, WA 98104 USA. Contact him at joseph.williams@infosys.com.

LISA NEE is the privacy engineering lead with Infosys, Osprey, FL 34229 USA. Contact her at lisa.nee@infosys.com.