A PETs Primer- What Are They, Why Are They Important, And What You Should You Know

Introduction

The digital advertising industry faces an array of compliance challenges when engaged in the collection, use, and sharing of consumer data. Ranging from the delivery and targeting of ads, to measuring and reporting on ad campaign performance, to detecting ad fraud and viewability, the new normal is framed by constant change. In the United States, that change is highlighted by expanding state privacy legislation paired with the increased focus by regulators at the national level such as the Federal Trade Commission (FTC) on the collection of sensitive data for digital advertising and the constant flood of class action litigation over the use of pixels and other online trackers. In Europe, there is continued active enforcement of the General Data Protection Regulation (GDPR) against companies in the digital advertising ecosystem (and the large fines attached to those enforcement actions), along with a new round of EU data tech laws. And, of course, there is the anticipated demise of third-party cookies. As a result, advertisers, publishers, platforms, and others in the ad tech space are constantly searching for tools to help address these challenges while effectively operating a data-driven business in a compliant way.

Chief among these tools are privacy enhancing technologies (PETs). PETs are not new and, in fact, have been around for quite some time. However, there is an increased focus on using PETs in the digital advertising space in light of privacy challenges. This edition of Practical Privacy Insights summarizes how organizations can use PETs to address privacy compliance obligations, discusses some of the risks and pitfalls related to the use of PETs, explores what regulators are thinking and saying about the use of PETs, and provides a practical primer on how to address PETs in your business and what to expect in the coming year.

What are PETs?

PETs are an umbrella term for specific technologies that allow for the processing and usage of data while simultaneously preserving the privacy of the data subjects. PETs maintain the informational value of the data with varying accuracy while providing different ways to enhance privacy protections. PETs cover a range of different technologies, but all involve balancing utility with privacy. The goal is to maintain the privacy of data subjects while still enabling the use of data.

There is still some relative inconsistency within the industry in connection with how these various technologies are defined and categorized. For example, the Centre of Information Policy Leadership (CIPL) puts PETs in three categories: cryptographic tools, distributed analytics tools, and tools for pseudonymization and anonymization.[1] The Organization for Economic Cooperation and Development (OECD) categorizes PETs into one of four buckets: data obfuscation tools, encrypted data processing tools, federated and distributed analytics, and data accountability tools.[2] With the exception of the data accountability category (which OECD notes are not considered as PETs in the strict sense).[3] the OECD categories largely track the CIPL categories. But the way each technical solution is categorized may be slightly different, e.g., OECD categorizes zero-knowledge proof as a data obfuscation tool while CIPL puts it in the cryptographic category.[4] Unlike CIPL and OECD, the Information Commissioner's

[1] Centre for Information Policy Leadership, Privacy-Enhancing and Privacy-Preserving Technologies: Understanding the Role of PETs and PPTs in the Digital Age, at 9 (2023) ("CIPL PETs Report").

[2] The Organization for Economic Cooperation and Development, Emerging Privacy Enhancing Technologies: Current Regulatory and Policy Approaches, No. 351 at 4-5 (Mar. 2023) ("OECD PETs Report").

[3] OECD PETs Report at 23.

[4] OECD Pets Report at 4-5; CIPL PETs Report at 9.

Office (ICO) categorizes PETs as either those that provide input privacy or those that provide output privacy.[5] As there appears to be some lack of complete alignment on standard definitions and taxonomy for PETs, you may find that certain vendor and business partner definitions and categorizations vary slightly, and certain solutions may fit into multiple buckets. Notably, one of the goals of the IAB Tech Lab's PET Initiative is to help standardize these definitions, categorizations, and use cases.[6] But for now when working on any data project involving PETs it is best practice to ensure that everyone involved is on the same page.

Examples of some emerging PETs which are showing promise in the digital advertising space including the following:

Homomorphic Encryption:

Homomorphic encryption allows computations on encrypted data without revealing the data to the processing party. This differs from standard data processing methods which typically call for the data to be visible to the processing party (i.e., the data must be unencrypted prior to processing, and the operations are conducted on data in the clear). Conducting processing operations on encrypted data is technically complex and generally requires significant computing power. However, algorithmic and hardware improvements, in addition to open-source compilers and libraries, are making these solutions easier to implement.

Federated Learning and Distributed Analytics:

Solutions using federated learning and distributed analytics allow for insights and information to be gained from data without the need to share or disclose the data. These solutions typically involve some collection and preprocessing of data which is done at the data source (e.g., on a data subject's own device or within a data contributor's own systems) and then transferring only summary results or statistics devoid of personal data to be further analyzed, combined, and/or synthesized with other results. One popular example of the use of such systems is the training of speech recognition models used in popular digital assistants wherein a local model is trained on a user's device against locally stored user data, and the model updates (but not the underlying user data) are sent to a centralized server where they are combined with updates from other users' devices and are then used to update a single global model.

Trusted Execution Environments (TEEs):

TEEs are secure areas of a central processing unit (CPU). There, information is stored, accessed, and code is run in such a way that it is completely isolated from the rest of the system. This means that processing is limited to a specific part of the CPU. The TEE can access outside information, but any sources outside the TEE cannot access information in the TEE.

Synthetic Data:

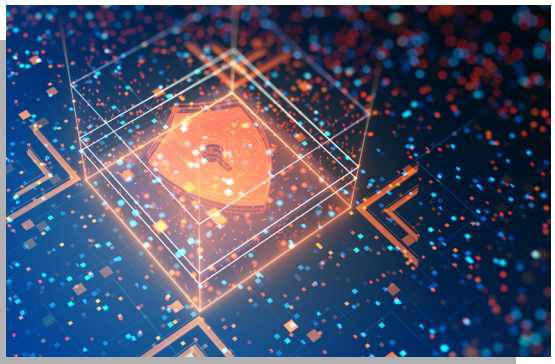
Synthetic data solutions involve exactly what the name suggests - the generation of artificial data attributes - disconnected from the real-world - from real data sets, wherein the generated synthetic data reproduces the characteristics and structure of the real data. Data synthesis algorithms can be used to generate the synthetic data. Financial institutions currently use synthetic data to train fraud detection models without putting sensitive data at risk.[7] Fraud detection models require a large amount of data for training. Synthetic data solutions take examples of real-world fraud, which can include individuals' financial information and unique identifiers, and generate larger sets of artificial data for training that contain similar characteristics. This PET provides an alternative to use of real data.

---

[5] Information Commissioner's Office, Privacy-Enhancing technologies , at 6-7 (June 2023) ("ICO PETs Report").

[6] See Privacy Enhancing Technologies (PETs) Initiative, https://iabtechlab.com/pets/

[7] Gonçalo Ribero, Synthetic Data Applications in Finance, FORBES (Apr. 3, 2024) https://www.forbes.com/sites/forbestechcouncil/2024/04/03/synthetic-data-applications-in-finance/?sh=7808e48d695d



Differential Privacy:

In solutions implementing differential privacy, minor changes are made to the data ("noise" is added) under the theory that slight changes can help de-identify the inputs without significantly impacting the results of the processing. This involves a careful balancing between privacy/security and utility. Add too much noise and the results will not be accurate, but too little noise could lead to potential re-identification.

Secure Multi-Party Computations (SMPC or MPC):

This solution involves using a set of tools or protocols that allow the joint processing of data without the parties sharing all their data. Instead, parties divide their data into distinct parts, which are then distributed amongst the different parties. This means that divided data, which is held across multiple parties, can be analyzed without transferring all the data. Each party runs computations on the parts they have been given before distributing the results among the other parties. The only information that is shared is the result of the computation using a cryptographic technique called secret sharing.

Private Set Intersection (PSI):

This solution allows two parties to find where their data overlaps without revealing or sharing their data sets. PSIs can also compute the size of the overlap and its aggregate statistics.

Zero-knowledge Proof (ZKP):

ZKPs allow for a party to verify knowledge without revealing the underlying data. For example, an advertiser can use ZKP to represent to a third party that a dataset contains information regarding middle-aged women interested in biking without revealing any of the actual data. ZKPs can also be used to confirm an individual is over 18 without revealing their birthdate, or that someone owns an asset without revealing past transactions.[8] The key differentiator of ZKPs from other PETs is that the data itself, even the results, are not shared. Instead, parties are able to prove that certain properties exist within the dataset. For this reason, ZKPs are increasingly considered for the implementation of data identity management systems, like the EU's proposed digital identity wallets.[9]

How Can PETs Help?

In broad strokes, PETs can assist organizations in complying with data privacy principles and various legal obligations by, among other things:

- minimizing the amount of data to be collected, shared, and processed;
- helping to meet requirements or standards related to pseudonymization and de-identification;
- enhancing security;
- ensuring data is only used for specific, documented use purposes;
- helping to address challenges related to cross-border data transfers; and
- improving the ability to audit data processing practices.

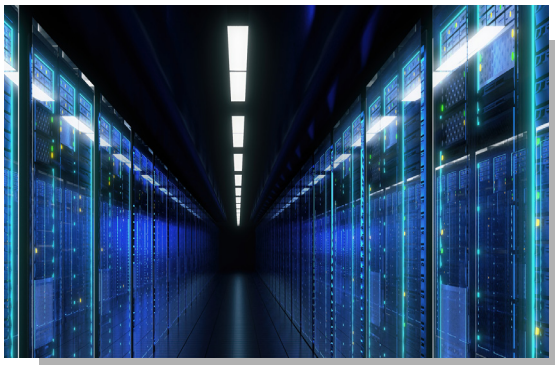
[8] Information Commissioner's Office, Chapter 5: Privacy-enhancing technologies (PETs), Draft Guidance at 29 (Sept. 2022) https://ico.org.uk/media/about-the-ico/consultations/4021464/chapter-5-anonymisation-pets.pdf.

[9] OECD PETs Report at 17.

PETs can also play a role in enforcing privacy policies and ensuring data use cases are aligned with privacy promises made to consumers. And PETs can be incorporated into privacy by design principles at the initial stages and throughout the development of new products and services.[10]

For example, in the digital advertising space, federated learning may be used for attributing ad interactions to conversion data for an individual on the individual's own cell phone or other device, and then sharing only the aggregated summary attribution results. Google's Topics API is a helpful illustration of a federated learning PET. It uses browser history to predict a user's interests, but limits the processing of personal information on the user's device, and only shares high level summary topics of interests with advertisers.[11] Similarly, Google's Protected Audiences API allows audience data to be stored and analyzed on a user's browser during an ad auction that is run locally within the user's browser.[12]

All these solutions help minimize the amount of personal data collected, used, and shared. Importantly, data minimization principles are codified into many privacy laws and regulations. The California Consumer Privacy Act (CCPA) requires that "a business' collection, use, retention, and sharing of a consumer's personal information shall be reasonably necessary and proportionate to achieve the purposes for which the personal information was collected or processed,"[13] whereas the GDPR provides that personal data shall be "limited to what is necessary in relation to the purposes for which they are processed."[14] By their nature and purpose, federated learning and distributed analytics are helpful tools to consider in complying with those data minimization requirements. These solutions could mitigate the burden in responding to data subject requests and obligations to secure or delete personal data. Along those lines, federated learning can help comply with data localization requirements or obligations related to international data transfers by eliminating the need to transfer personal data across borders. Lastly, federated learning can assist organizations ensure they do not have data in their possession that could be inadvertently used for purposes beyond the purposes that data was collected for, or exceed the notice provided to consumers.



Other solutions, like differential privacy, synthetic data, homomorphic encryption, and private set intersection using encryption, allow data to be shared and processed in a modified, encrypted, or obfuscated form. For example, synthetic data can be shared with partners rather than the real, original data set to create and train a model used for ad targeting or measurement. Or private set intersection may be used to match encrypted advertiser and publisher data sets for targeting purposes or to calculate campaign conversion rates without sharing the underlying personal information.

In an ideal situation, these solutions would sufficiently transform the data to the point where the data falls outside the scope of privacy laws. In that case, many of the same compliance benefits described above in connection with federated learning and distributed analytics would similarly apply to these solutions. However, laws vary in terms of requirements for data to be sufficiently de-identified such that it is not considered personal data. This will depend on factors like whether it is possible to re-identify or link the information to individuals, and, if so, how difficult it would be to do so. Notably, data which is merely pseudonymized is generally considered

[10] OECD PETs Report at 8-9.

[11] See Google Privacy Sandbox, Topics API for Web Overview, https://developers.google.com/privacy-sandbox/relevance/topics#epoch.

[12] See Google Privacy Sandbox, Protected Audience API Overview, https://developers.google.com/privacy-sandbox/relevance/protected-audience#what.

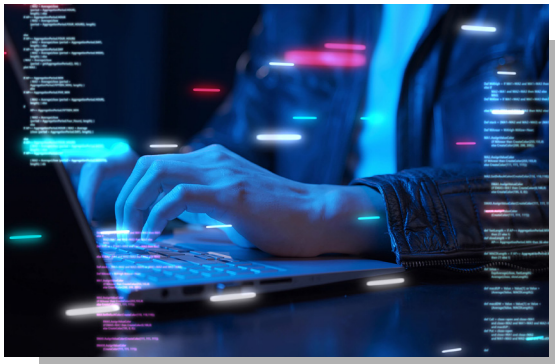
[13] Cal. Civ. Code § 1798.100(c).

[14] GDPR, Article 5 (1)(c).

personal data under applicable laws because it can still be used to identify individuals especially when combined with other identifiable data.[15] And states have taken a mixed approach to consumer privacy rights for pseudonymized data. Some states have exceptions for all consumer rights with respect to pseudonymized data, while others only require opt-outs for pseudonymized data, and still others do not explicitly include exemptions for pseudonymized data at all.[16] So, determining the legal impact of the use of PETs involving modified, encrypted, or obfuscated data requires a careful analysis of how applicable laws define and treat data which is anonymized, deidentified, or pseudonymized.

However, even if these solutions do not meet anonymization or de-identification privacy law requirements, these PETs can still be useful security measures. Separately, many data protection laws include specific data security requirements and require the use of appropriate technical and organizational security measures to protect data.[17] Use of encrypted data and encrypted data processing techniques may help organizations comply with those data security requirements or data security promises they have made to consumers. In addition, many data laws provide explicit exemptions from obligations and safe harbors for encrypted data (e.g., data breach notification requirements).[18]

The foregoing highlights only a few different use cases for PETs. It is important to keep in mind that PETs are often complementary, and several different types of PETs may be used together to achieve additional benefits. And, although this paper primarily focuses on use of PETs in connection with their privacy protection capabilities, organizations can use PETs to protect the commercial or business interests of organizations, e.g., by helping to ensure that vendors and partners comply with contractual restrictions related to use and processing of an organization's data.



What Have the Regulators Said about PETs?

Regulators understand the potential of PETs and the role they can play in protecting and preserving privacy and in complying with privacy obligations. In a 2021 blog article, for example, the Office of the Privacy Commissioner of Canada (OPC) noted that both federated learning and differential privacy offer promising approaches for businesses looking to ensure data remains private and secure.[19] Regulators and governments have also established contests and sandboxes to promote and encourage the development of PET technology.[20]

Regulators have also expressed concerns about the level of maturity, technical knowledge, and experience required for proper implementation, and about misrepresenting the capabilities of PETs. For example, in its recent guidance related to use of PETs, the ICO noted that "some PETs may not be sufficiently mature in terms of their scalability, availability of standards and their robustness to attacks.[21] The ICO also stated that "PETs can require significant expertise to set up and use appropriately," and cautioned: "if you do not have required expertise, then you should consider using an off-the-shelf product or service

[15] See Fernando Bohorquez, Jr., Gerald Ferguson, et al., The Post-Cookie Digital Advertising Landscape: Planning for Privacy Compliance in Unsettled Terrain, (May 16, 2023) https://admin.adventures-in-law.com/wp-content/uploads/sites/2/2023/07/The-Post-Cookie-Digital-Advertising-Landscape-Planning-for-Privacy-Compliance-in-Unsettled-Terrain-w-038-4579.pdf ("Bohorquez/Ferguson, Post-Cookie Privacy Compliance ").

[16] Katherine Danko, Comparing U.S. Comprehensive State Privacy Laws: Treatment of Pseudonymous Data, NAI (Aug. 14, 2023) https://thenai.org/comparing-u-s-comprehensive-state-privacy-laws-treatment-of-pseudonymous-data/.

[17] See 15 U.S.C. § 6502 et seq.

[18] See Iowa Code § 715C.1 et seq.

[19] Office of the Privacy Commissioner of Canada, Privacy Tech-Know blog: Privacy Enhancing Technologies for Businesses (Apr.12, 2021), https://www.priv.gc.ca/en/blog/20210412/.

[20] Privacy-Enhancing Technologies: Privacy Challenges, U.K.-U.S. Privacy Challenges, https://petsprizechallenges.com/.

[21] ICO PETs Report at 6.

that provides an appropriate level of support.[22] In a recent post, the FTC cautioned companies about misrepresenting the capabilities of PETs, stating "companies building products that use MPC, oblivious proxies (technologies which rely on trusted third-party intermediaries to strip personal data from information exchanged between parties before forwarding on the remaining data), or other PETs still need to keep Section 5 in mind and ensure they're not engaging in unfair or deceptive acts or practices," and reminding companies that it "has brought cases against companies that claimed specific technology-based security or privacy guarantees that they allegedly failed to provide."[23]

How Should You Approach Use of PETs?

Notwithstanding that some of these technologies are still evolving, PETs are likely to become increasingly relevant in the coming years. The following brief playbook on how to approach PETs can help the legal community when asked to opine on PETS.

1. Understand the Business Goals

Before embarking on any analysis of whether PETs may be the right tool to assist in accomplishing your team's business goals, you should make sure everyone is aligned on what those business goals are. At the end of the day, what is the organization trying to accomplish? Is the goal to develop a new product or service offering, to improve existing ones, to conduct some sort of internal measurement or analysis, or to monetize your own data by making it available to others? Taking time to make sure everyone is on the same page about what the goal is and what success looks like will pay dividends as you work through the process. Before you reach into your proverbial toolbox to find the right tool for the job you need to know what the job is.

2. Identify the Data and Data Flows

The next step is gaining a thorough understanding of what data is being leveraged and how it will be collected, shared, processed, and stored. Understanding what data is being leveraged will help determine whether such data triggers compliance obligations under the various privacy laws, rules, and regulations, or data security concerns. The type of personal information—e.g., emails, IP addresses, device IDs, location data, SSN, etc.—will invariably inform what PET or combination of PETs are best suited to meet both business and privacy goals. Similarly, determining where the data will be stored, e.g., on premises, on device, or with a third party, and whether and how the data will be transferred between the different parties involved, will assist in steering you to what PETs would be most on point. For example, federated learning and distributed analytics solutions may enable the business to accomplish its objectives using data stored locally on a user's device, and for use cases which require the sharing or transferring of data between parties PETs that implement encryption and obfuscation may be good options. Finally, you need to get a firm grip on the output data, to the extent applicable. Will the results of any processing include personal or pseudonymized data? If not, will the output data be aggregated or anonymized in some way? And, if so, how? Only by fully understanding the data and data flows at play, and the business goals at heart, will you be able to begin to analyze the various privacy and security concerns, and what specific tools would be helpful in mitigating those concerns.

3. Analyze the Applicable Laws, Regulatory Risk, and Other Concerns to be Addressed

Once you have a good understanding of the business goals, the use cases, and have identified the data and data flows involved, the next step is to assess the applicable legal obligations, regulatory risks, and other concerns that need to be

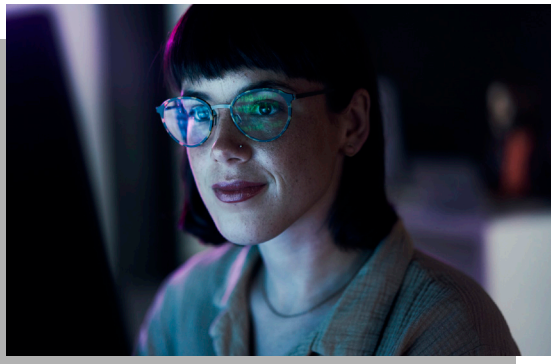
---

[22] ICO PETs Report at 6.

[23] Simon Fondrie-Teitler, Federal Trade Commission Office of Technology, Keeping Your Privacy Enhancing Technology (PET) Promises, (Feb. 1, 2024) https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2024/02/keeping-your-privacy-enhancing-technology-pet-promises.

addressed. The lodestar here–like in almost all situations–is a risk assessment. The departure point in analyzing the legal and regulatory risk is identifying the applicable legal frameworks. The information gathered in steps 1 and 2 above will help you determine what legal frameworks apply (e.g., U.S. state privacy laws, GDPR, etc.), and that will in turn assist in determining all of the various legal obligations that are triggered, including, e.g., required privacy notices, requirements to obtain consent from data subjects for use of their data, or to provide them with an opportunity to opt-out of such use, and requirements related to facilitating other data subject rights such as access and deletion. But the analysis should go beyond understanding the requirements triggered by specific privacy and data security statutes. For example, you should also consider obligations prompted by any representations you have made to consumers and other data subjects, any contractual obligations you may have in connection with agreements you have entered, and risks related to potential reputational harm in connection with your use or sharing of data.

4. Determine and Define the PETs to Address the Goals



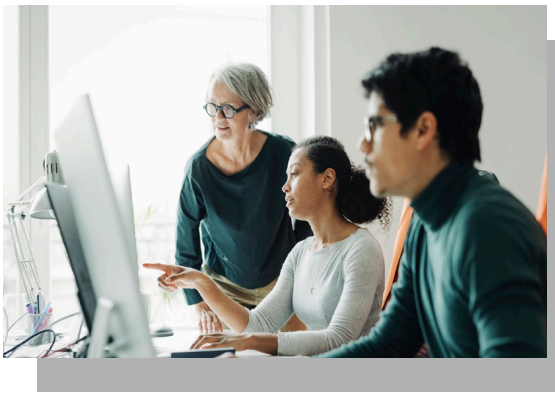
As summarized above, PETs can be broken down into three groups: those that help encrypt data (e.g., homomorphic encryption), PETs that limit or control access to data (e.g., federated learning), and those that obfuscate data (e.g., differential privacy and synthetic data). What is important to understand, and is frankly quite complex, is that various kinds of PETs may have similar underlying goals while being nuanced in their implementation and results. For instance, zero knowledge proofs, multi-party computations, and federated learning can all be used to further data minimization and purpose limitations, but they do so in quite different ways. Moreover, and more importantly, no one PET is necessarily mutually exclusive to the other, and they can be used together to achieve a wide variety of privacy goals. That is why PETs are frequently used in combination with each other. For example, homomorphic encryption can be used with federated learning to hide the machine learning models before being sent to the global model update shared by the parties, or it can be combined with multi-party computing to encrypt each party's private data inputs. You need to fully define the PETs being deployed to ensure they can be leveraged to complement one another and that working together they fill the problematic gaps. In addition, clearly defining the PETs being used to address each goal will help avoid the pitfall of confusing or conflating the use of one PET with the actual use of many.

5. Be Mindful of Common PET Pitfalls and Challenges

As you work through this analysis, it may be helpful to keep some of these common PET pitfalls in mind:

The silver bullet. One of the challenges with PETs is that some people perceive them as a privacy cure-all for any new product or service. Obviously, that is not the case. But to a layperson, in particular, it may not be so obvious that implementing PETs in a new business line or new venture will not automatically address all privacy issues. As with anything else, it is essential to understand the core of what the business is trying to achieve and what privacy goals PETs are intended to meet, and to do a thorough analysis to determine whether a certain PET technology (or, more likely, a combination of different technologies) can help address and mitigate privacy concerns. Rather than looking at PETs as being the one single solution to solve your privacy woes, they should be considered a set of tools that can help you address specific legal, regulatory, and business concerns. For example, if the key business goal is to enrich current data sets by combining data with other parties in a data clean room, is the team most interested in leveraging PETs to ensure data security, data minimization, or data anonymization?[24] Is the team mostconcerned with controlling the access to the data from the other parties contributing to the data clean room? Or is the key concern to maintain the security of the personal data at all inputs given its sensitivity? The answers to these kinds of questions will help determine what PETs you should guide the team in considering.

[24] IAB Technology Laboratory, Data Clean Rooms: Guidance and Recommended Practices, Version 1.0 (July 5, 2023) https://iabtechlab.com/datacleanrooms/.



Don't over-promise. As mentioned above, regulators are going to be carefully reviewing the promises business are making to consumers around their use of PETs and, as with anything else, business should be careful to ensure that the representations they make to consumers are truthful and accurate. Given the technical complexity surrounding some PET implementations and the risk of improper implementation organizations thoroughly vet the statements they make to consumers about how their use of these technologies work and how they protect data subject privacy and security.

The specter of re-identification also looms large. One of the purported key benefits of many PETs is that they anonymize, de-identify, or make pseudonymous data under applicable privacy laws such that the data may no longer constitute personal data or it meets certain exemption criteria. But the GDPR, U.S. state laws like the CCPA, and most other privacy laws have strict requirements for when data qualifies as being anonymized or de-identified, and where exceptions apply for pseudonymized data, there are similar requirements and qualifications.[25] That bar may be high and difficult to meet. If, for example, you are relying on PETs like differential privacy or use of synthetic data to help anonymize or de-identify data, you need to make sure that the random noise and analytic outputs prevent the re-identification of any individual data and meet the strict requirements under applicable laws. Part of that analysis should consider the extent to which any summary data or centralized model created is susceptible to data leakage, model inversion, membership inference, or other techniques that could be used to re-identify individuals.

Authors

Fernando Bohorquez
Partner
BakerHostetler

Aaron Goodman
Counsel
BakerHostetler

Kathryn Prince
Associate
BakerHostetler

[25] See Bohorquez/Ferguson, Post-Cookie Privacy Compliance.