ISSUE BRIEF U.S. Policy

Privacy Enhancing Technologies for EdTech Service Providers

May 2026

FUTURE OF PRIVACY FORUM

AUTHOR

Jim Siegl, Senior Fellow, FPF

About This Document

This document provides a strategic guide for EdTech providers on implementing Privacy Enhancing Technologies (PETs) to protect student data during research and product development. It details specific methods such as synthetic data, differential privacy, and homomorphic encryption, explaining how these tools allow vendors to gain valuable insights while minimizing the risk of re-identifying individual students. Ultimately, the report emphasizes that while PETs provide powerful tools for reducing data exposure, the decision on whether to apply them should be accompanied by an understanding of their impacts on analytical accuracy and integrated into a broader framework of transparent data governance and contractual obligations.

Primary audiences: Education Technology Service Providers

The Future of Privacy Forum (FPF) is a non-profit organization that serves as a catalyst for privacy leadership and scholarship, advancing principled data practices in support of emerging technologies. Learn more about FPF by visiting fpf.org.

Privacy Enhancing Technologies (PETs) for EdTech vendors

Executive Summary

Educational research answers questions about program effectiveness, equity, access, and long-term outcomes. EdTech vendors typically need student-level data and enough context to ensure their research is effective and accomplishes their goal.[1] At the same time, disclosing student data to EdTech vendors may increase the risk of reidentification. The risk is not limited to direct identifiers such as names or student IDs; uncommon or rare characteristics and small groups can also reveal too much.

Some analyses—particularly those that rely on predefined queries or aggregate statistics—can be completed without direct access to student-level data. Agencies can enable analysis through protected environments or Privacy-Enhancing Technologies (PETs). PETs are methods that help reduce the risks of sharing student data while also preserving its value. Importantly, PETs provide an additional layer of privacy and security for permitted uses of student data by EdTech vendors, such as using de-identified data for research or product improvement, within the bounds of applicable laws and vendor contracts.[2] While not every use of student data may require the use of PETs, they can be useful for lowering the amount of sensitive data shared, supporting safer publication of results, and protecting small groups. PETs can also support and complement good governance goals such as data minimization, role-based access controls, auditability, and enforceable agreements.

---

[1] Note: the types of student data an EdTech vendor may have access to can vary depending on the nature of the service provided and educational purpose fulfilled. Some vendors may already hold personally identifiable information or de-identified student data in the normal course of providing contracted educational services to a school that may be sufficient for certain secondary uses (when properly de-identified), while other vendors may only have access to limited or aggregate data sets to fulfill the educational purpose. Based on the nature of the vendor relationship with an educational institution, uses of student data for purposes other than directly providing the educational service, such as for research or product improvement and testing, may subsequently require additional de-identified data disclosures, as permitted under applicable law or contract.

[2] Note: many state student privacy laws and vendor contracts, such as the National Data Privacy Agreement (NDPA) model contract developed by the Student Data Privacy Consortium, allow research and product development using de-identified student data. The specific allowances for a vendor's use of de-identified student data for purposes of research or product development depends on the specific state and vendor contracting terms involved in a specific use case. The US Department of Education's Privacy Technical Assistance Center (PTAC) has also issued guidance on uses of de-identified student data for research and product improvement—see its Model Terms of Service, first released in 2014 and revised in 2016.

How PETs Can Help EdTech vendors

PETs can be applied to reduce the amount of sensitive data shared. Instead of sending full datasets, PETs can allow computation where the data already resides or within a secure enclave. PETs can also allow you to publish findings with less risk of identifying students**.** Techniques such as differential privacy can add “noise” to outputs, making it harder to reidentify students with rare or unique characteristics.

Protecting small groups when reporting is important to ensure data stays protected and cannot be linked back to any specific student. When subgroup sizes are small, PETs can reduce the chance that tables, dashboards, or any other outputs inadvertently reveal outcomes for a single student. It’s also necessary to protect sensitive attributes**.** PETs can help minimize exposure of highly sensitive fields by supporting computations that do not require those attributes to be accessible outside a controlled environment.

Deciding What PETs to Use

PETs reduce exposure, but each one imposes a cost on analytical fidelity or operational flexibility. For EdTech vendors, that cost shows up differently depending on the workflow. Four practical factors help to guide you in deciding which PET should be used:

- **Business/product purpose:** What is the vendor trying to accomplish — system testing, staff training, product analytics or improvement, or collaborative research with a researcher or education agency? The answer determines which PET is appropriate and what analytical fidelity is required.
- **Data sensitivity and contractual obligations :** Consider what student data the vendor has access to under its agreements with schools or agencies, including restrictions on secondary use, re-disclosure, and retention. PET selection should reduce exposure to align with those obligations, not just meet a technical threshold.
- **Operational need:** Identify where in the vendor's workflow privacy risk is highest — data ingestion, model training, reporting outputs, or vendor-to-agency data returns — and match PET selection to that pressure point.
- **Privacy-accuracy tradeoff**: Different vendor workflows tolerate different levels of analytical imprecision. A system integration test using synthetic data doesn't need exact fidelity. A product effectiveness dashboard shared with a district does. Identifying the acceptable tradeoff early determines which PET is viable.

Synthetic Data

EdTech vendors need student data for things like the training and onboarding of employees or running simulations on their product. However, using real student records for these purposes does not keep that information private or secure. In situations like these, synthetic data serves a valuable purpose.

Synthetic data is artificially generated data that **mimics the statistical properties** of real-world datasets without containing any records from actual individuals. By preserving the **distributions, correlations, and patterns** in the original data, synthetic sets enable robust testing and analysis while significantly mitigating the risk of re-identification.

Synthetic extracts are particularly useful when teams need realistic test data but do not need real identities or exact counts for production decisions. Synthetic data preserves data architecture and statistical patterns well enough for system testing and staff training, but may not accurately represent rare student populations or edge cases that surface in production. A vendor that trains staff on synthetic data reflecting average district demographics may inadequately prepare staff for smaller or more diverse districts.

However, synthetic data does not automatically provide formal privacy guarantees, and its analytical validity depends on the generation approach. It is generally less suitable for analyses that require precise estimates, rare populations, or exact counts, and should be evaluated for both privacy risk and statistical fidelity before use.

Differential Privacy

When exact student level data is not required, EdTech vendors can use **differential privacy** (DP) to do a multitude of things, including training staff, evaluating their product, or sharing aggregate information from student data without exposing information about any individual student. EdTech vendors are able to create realistic instructional materials, dashboards, or reporting examples without relying on exact data tied to real students.

DP is a formal privacy framework that bounds how much the inclusion of any one student's data can influence reported results. It does this by introducing carefully calibrated statistical noise, so that outputs are similar whether or not a particular student's record is included, while still preserving overall patterns in the data. The strength of this protection is governed by a parameter (often called ε, or "epsilon") that reflects a tradeoff between privacy and precision.

DP is well-suited for aggregate reporting—usage dashboards, outcome summaries, product effectiveness metrics—but introduces noise that grows more distorting as the population being analyzed gets smaller. A vendor reporting on a small school's reading intervention outcomes

under DP may produce results that are misleading at that scale, even if the privacy guarantee is technically correct.

Homomorphic Encryption

There are times when an EdTech vendor cannot have access to raw student data, however, they still need to do analyses on that student data on behalf of a state or district. In these situations, **homomorphic encryption (HE)** can help by allowing the EdTech vendor to analyze the data while it's encrypted so that computations can take place on the data while it also remains protected.

This works by the state or district encrypting data prior to sending it to the EdTech vendor. Once received, the EdTech vendor runs the agreed-upon analytics on the encrypted data without ever seeing the underlying raw student data. Once the analytics are completed by the EdTech vendor, those encrypted results are returned back to the state or district, who decrypts the outputs internally.

For example, an EdTech vendor could help an agency to analyze program effectiveness across districts and student groups without ever being exposed to the agency's individual-level data. This approach reduces data exposure by preventing the EdTech vendor from viewing or retaining student-level information, while supporting governance principles such as least privilege and purpose limitation.

However, this approach does not eliminate all risk. Aggregated outputs may still reveal sensitive information about small groups, making disclosure controls essential. In addition, homomorphic encryption introduces several practical limitations. It is computationally intensive and can significantly increase processing time and cost compared to standard analysis, particularly for large datasets or complex models. Not all types of analyses are feasible under current HE schemes; operations such as complex joins, iterative algorithms, or certain machine learning methods may be limited or require substantial adaptation.

There are also implementation and governance considerations. Incorrect implementation or parameter choices can weaken security guarantees, and key management becomes critical, as loss or compromise of the decryption key would prevent access to results or expose sensitive data. While HE protects data during computation, it does not inherently protect against all forms of inference; repeated queries or differencing attacks on outputs may still allow sensitive information to be inferred if disclosure controls are not carefully applied.

Conclusion

EdTech vendors do not have to choose between research data fidelity and student privacy. With the right approach, both goals can be achieved simultaneously because PETs offer

mathematically verifiable tools for conducting studies that minimize data exposure without compromising statistical validity. In practice, this means selecting approaches that align with the research question, the data's sensitivity, and the acceptable level of precision—recognizing that different PETs introduce tradeoffs among utility, complexity, and risk.

At the same time, PETs do not substitute for the need for disclosure review or governance. EdTech vendors operate under a dual obligation: to deliver useful analytics and product capabilities, and to handle student data in ways that honor the trust schools and agencies have placed in them. PETs can help meet both obligations simultaneously, but only when the tradeoffs are understood and disclosed.

A vendor that deploys synthetic data for system testing without validating fidelity, or that generates DP-protected dashboards without disclosing noise parameters, has technically implemented a PET without actually managing privacy risk responsibly. Even when not required by laws or regulations, PETs can help EdTech providers to build products and analytical workflows in ways that that agency customers can trust and hold up when a district asks how the numbers were produced.

PETs are most effective for EdTech vendors when they are matched to specific workflow needs, integrated into data governance documentation and contractual representations, and disclosed transparently to agency partners. Used that way, they reduce exposure, support more flexible data relationships, and strengthen the vendor's long-term credibility with the institutions it serves.

Washington, DC | Brussels | Singapore

FPF.org