Comparison of Privacy Enhancing Technologies (PETs)

PET	Approach	Benefits	Limitations	Example	Data Utility	Complexity
Differential Privacy	A method of data anonymization that injects "noise" to prevent the identification of sensitive individual data. The output is nearly the same whether or not a student's record is included.	Differential privacy is most effective when you are looking for patterns, trends, or statistics across a large group. A better alternative to simply removing names, which can re-identify students with other available information.	Can distort results when there are small groups, making data less reliable. If you need 100% precision at the row level, DP is not the right tool.	Differential privacy is commonly used in large data sets Public reporting, dashboards, and large-scale aggregate statistics.	**High (Aggregates)**: Maintains statistical trends but adds noise to protect individuals.	**Medium/High**: Requires careful calibration of the "privacy budget" and noise levels.
Federated Learning[1]	Each school/district uses its own data in an AI model, and the data remains within the school/district environment. The updates are sent out and combined with other school/district's updates for an improved shared model that is sent back out to the school/district.	Less sharing of raw data (and more institutional control), but still able to support collaboration with other schools/districts without pooling data. Better model performance than with one school's training alone, improving accuracy.	Model updates can leak information, revealing details about the data, so you often need to use it in conjunction with another PET. FL is slow, making it a poor choice if you need a model updated in real time.	Developing an at-risk model to predict dropouts or course failure.	**High**: Models are trained on raw data locally, preserving accuracy without pooling records.	**High**: Requires standardized data models and coordinated infrastructure across agencies.
Homomorphic Encryption	A method enabling encrypted computations to be conducted on encrypted data without needing to	Districts can collaborate while sharing less readable data. Reduced breach impact because attackers	For day-to-day use, it can be expensive and impractical. Decryption key management is imperative to ensure	Encrypting attendance data and sending it to a vendor to compute dropout/failure risk.	**Very High**: Computations on encrypted data yield results identical to those on raw data.	**Very High**: High computational cost; often slower than traditional processing.

[1] *Federated Learning is technically neither anonymization nor de-identification. It is better described as Data Minimization. It is often used alongside other PETs, such as DP.*

PET	Approach	Benefits	Limitations	Example	Data Utility	Complexity
	decrypt it first. The decrypted results match what would have been obtained by performing the calculations on the original unencrypted data.	can only obtain encrypted data and encrypted outputs.	you can read results while keeping student data private.			
Secure Multiparty Computation	A technique allowing multiple parties to process their combined data without any party needing to share all its information with the others. This approach minimizes the risk of exposing sensitive information.	Multiple institutions can jointly run analyses on combined populations and get exact results without handing over raw datasets because no one entity holds all the data.	Operationally, all parties must agree on computations and managing updates, which can be difficult when dealing with different policies. Results can still leak sensitive data when handling granular information, necessitating the use of additional privacy controls.	Determining scholarships or financial aid when needing access to data from multiple sources (financial aid office, scholarship organizations, government agencies, etc.)	**Very High**: Produces exact mathematical results without any party seeing the input data.	**Very High**: Significant computational and network overhead; requires specialized protocols.
Synthetic Data	A model is trained on real student data, learns overall patterns, and generates new synthetic data that shows the same trends without containing any real student records.	Can share realistic data without sharing individual student information. Easier and faster access to data since you aren't relying on real data to be obtained.	Trade-off between accuracy and privacy, with more accurate synthetic data potentially leaking sensitive information. Requires expertise to test data and verify it is safe to use.	Software testing, system integration, and exploratory data analysis without PII.	**Medium**: Mimics statistical patterns but does not represent actual individuals.	**Medium**: Requires training a model on real data to generate the artificial records.
Trusted Execution Environment	Secure virtual computing spaces that enable the execution of code and access to data in an	Allows for more secure collaboration and multiparty analytics without needing to share	Setting up the enclave can be expensive and complex.	Secure "enclaves" for researchers to run approved code on sensitive datasets.	**Very High**: Allows computation on raw data within an isolated,	**Medium**: Depends on specific hardware availability and secure cloud configurations.

PET	Approach	Benefits	Limitations	Example	Data Utility	Complexity
	isolated manner, detached from the rest of the system. This isolated processing protects against unauthorized access, ensuring the confidentiality and integrity of sensitive data, 6 also known as a secure enclave.	datasets with the other party. Strong privacy for student information and better adherence to compliance standards.	Outputs need to be carefully designed to prevent sensitive information from leaking.		hardware-protected space.	