Privacy-preserving machine learning transforms how educational AI handles student data

The convergence of federated learning, differential privacy, and regulatory compliance frameworks now enables adaptive learning systems to deliver personalized education while protecting student privacy with mathematical guarantees. Federated Deep Knowledge Tracing (FDKT), introduced by Wu et al. in 2021, ACM Digital Library demonstrates that **knowledge tracing models can achieve comparable accuracy to centralized approaches while keeping sensitive behavioral data distributed across educational institutions**. Meanwhile, differential privacy mechanisms calibrated for educational contexts—with epsilon values between 1-10 for most applications—can protect individual students while preserving analytical utility. EdTech platforms operating across US and EU jurisdictions must navigate both FERPA's "school official" exception framework and GDPR's explicit consent requirements for children, with the EU AI Act now classifying educational AI as high-risk, Feedbackfruits triggering mandatory compliance requirements by 2026-2027.

Federated learning architectures enable cross-institutional knowledge tracing without data centralization

The foundational work on Federated Deep Knowledge Tracing by Wu et al. (WSDM 2021) ACM Digital Library addresses three critical challenges in educational data: scarcity across institutional silos, varying data quality from different learning schedules, and incomparability of knowledge states across different curricula. The architecture extends traditional LSTM-based Deep Knowledge Tracing to a federated setting where each educational institution maintains local models that are aggregated without sharing raw student interaction data.

Non-IID data distribution presents the most significant technical challenge for educational federated learning. Different schools have different curricula, student populations, and learning patterns—creating statistical heterogeneity that degrades standard FedAvg aggregation. arXiv Research from a 2025 Brazilian study simulating 50 schools with over 2 million student records found that federated DNNs achieved **61.23% accuracy compared to 63.96% for centralized XGBoost**—a 2.73 percentage point gap researchers deemed acceptable for LGPD compliance. arXiv arXiv

Alternative aggregation strategies show substantial improvements over FedAvg in heterogeneous educational settings:

- **FedProx** adds a proximal term constraining client model drift, arXiv achieving **22% average improvement** over FedAvg in heterogeneous settings GitHub +2 with tunable μ parameter (0.001-1 depending on dataset characteristics)
- **SCAFFOLD** uses control variates for variance reduction, arXiv arXiv achieving **89.1% accuracy** with fastest convergence (~70 rounds), Sciencexcel though it struggles with partial device participation arXiv

- **FedAtt** applies neural network attention mechanisms during aggregation, particularly effective for addressing demographic bias across institutions

Chu et al.'s attention-based personalized federated learning work (CIKM 2022) introduces personalized models for individual student subgroups derived from global models via meta-gradient updates, arXiv combined with self-supervised behavioral pretraining leveraging multiple modalities. Their Multi-Layer Personalized Federated Learning methodology optimizes inference accuracy across different layers—by course and by demographic subgroups—while mitigating bias in educational data availability.

For practical implementation, the **Flower framework** provides the most flexible open-source foundation, supporting PyTorch, TensorFlow, JAX, and scikit-learn Flower GitHub with built-in implementations of FedAvg, FedProx, FedYogi, and SCAFFOLD. arXiv Education-specific repositories include the federated-deep-knowledge-tracing implementation GitHub and the comprehensive pykt-toolkit covering DKT, SAKT, and SAINT models.

Differential privacy enables classroom analytics without individual identification

The LAK 2025 conference introduced **DEFLA** (Differential Privacy Framework for Learning Analytics) by Liu et al.—the first comprehensive differential privacy framework specifically designed for educational contexts. ACM Other conferences The framework provides a seven-step implementation process: problem assessment, threat modeling (using STRIDE framework), DP model selection, implementation, performance analysis, parameter tuning, and user communication.

Three distinct approaches apply differential privacy to educational ML pipelines, each with different privacy-utility tradeoffs:

Approach	Trust Model	Privacy Level	Utility Impact
Input Perturbation	Untrusted data collector	Strongest	Lowest utility loss in experiments
Objective Perturbation	Trusted training phase	Medium	Fluctuating utility
Prediction Perturbation (PATE)	Output-only protection	Weakest initial	Highest initial loss, improves with higher ε

Epsilon selection remains context-dependent, but emerging consensus from research and industry practice suggests:

- **$\varepsilon \leq 1$**: Strong formal guarantees for high-sensitivity applications
- **ε = 1-10**: Reasonable privacy-utility balance for most educational ML

- **Interactive dashboards**: 0.1-1 per query with 5-10 total per quarter
- **DP-SGD training**: ε = 1-10 with δ = 1e-6 for deep learning models

For knowledge tracing specifically, EDM 2025 research by Kabir et al. demonstrates that **DKT maintains 98-99% of non-private AUC at ε ≈ 7**. Their comparison shows DKT (non-private AUC: 0.80, private: 0.79), DKT+ (0.80 → 0.78), MonaCoBERT (0.79 → 0.75), and BKT (0.71 → 0.69)—confirming that simpler architectures preserve utility better under differential privacy.

Small cohort analytics presents particular challenges. DEFLA research explicitly notes that "DP struggles with small datasets (N<50), where its performance can degrade significantly." (ACM Other conferences) For classroom-level reporting with 25-30 students, recommended approaches include temporal aggregation across semesters, hierarchical reporting structures, suppression thresholds for counts below 10, and higher epsilon values (5-10) when necessary for utility.

Composition theorems are critical for continuous learning systems. Basic composition provides tight but conservative bounds (kε for k mechanisms), while advanced composition from Dwork, Rothblum, and Vadhan (2010) achieves √k improvement. For DP-SGD implementations, **Rényi DP accounting** through TensorFlow Privacy or Opacus provides the tightest bounds for Gaussian mechanisms.

FERPA defines education records broadly but leaves ML-derived profiles in ambiguous territory

Under 34 CFR § 99.3, **education records** include any records directly related to a student and maintained by an educational institution or its agents. (Splashtop) (Concentric AI) This definition clearly encompasses learning analytics outputs, clickstream data linked to individual students, knowledge tracing outputs tracking mastery progression, and predicted competencies when maintained by schools or their vendors.

The question of whether **ML model parameters** constitute education records lacks explicit DOE guidance. Legal scholars argue that models trained in ways that could potentially reconstruct or infer individual student information raise FERPA concerns, particularly when AI systems retain training data indefinitely contrary to FERPA's retention guidelines. (Flywire)

The **school official exception** (34 CFR § 99.31(a)(1)) provides the primary pathway for EdTech vendors to receive student PII without parental consent. Vendors must satisfy four requirements: performing an institutional service the school would otherwise use employees for, having legitimate educational interest in the records, operating under the school's direct control regarding data use and maintenance, and using records only for authorized purposes without re-disclosure.

The 2018 Agora Cyber Charter School decision established critical precedent: schools cannot require parents to waive FERPA rights as enrollment conditions, and Terms of Service allowing vendors to use data "for any purpose" are not FERPA-compliant. The Family Policy Compliance Office recommended use of Model Terms of Service (Hlerk) with explicit specifications of permitted uses.

De-identification under FERPA requires a case-by-case "reasonable determination" that student identity is not personally identifiable due to unique patterns, taking into account other reasonably available information. Unlike HIPAA's Safe Harbor, **FERPA provides no specific safe harbor listing required steps**. The Privacy Technical Assistance Center recommends techniques including blurring (reducing data precision), perturbation (small changes to prevent identification), and suppression (removing data from cells/rows)—while cautioning that behavioral patterns and learning trajectories can themselves be re-identifying.

A critical enforcement gap exists: **EdTech companies are NOT directly liable under FERPA**—only schools bear responsibility. This gap is increasingly filled by state laws including California's SOPIPA, Illinois's SOPPA, Colorado's Student Data Transparency Act, and New York Education Law §2-d, which establish direct vendor accountability with civil penalties. (Publicinterestprivacy)

GDPR requires purpose limitation and human oversight for educational AI affecting children

For EU schools, **public task** (Article 6(1)(e)) serves as the primary lawful basis for processing educational data, grounded in Education Act requirements across member states. However, EdTech providers acting as controllers cannot rely on public task (ico) and must typically establish legitimate interests through formal assessment or obtain explicit consent—complicated by the power imbalance between schools and students that may impair the "freely given" standard for valid consent.

Age of consent for information society services varies significantly across EU member states: 13 years in Belgium, Denmark, Finland, Spain, Sweden, and UK; 14 years in Austria, Bulgaria, Italy; 15 years in France and Greece; and 16 years in Germany, Ireland, Netherlands, and Romania. Parental consent verification methods range from email verification for low-risk cases to credit card verification, bank transfers, or video calls with trained staff for higher-risk processing.

Article 22's automated decision-making provisions apply when three cumulative conditions are met: the decision is based solely on automated processing, there is no meaningful human intervention, and the decision produces "legal effects" or "similarly significantly affects" the individual. For adaptive learning:

- University admissions, automated grade assignments, and progression-affecting rankings **likely trigger Article 22**
- Content recommendations and adaptive difficulty adjustments **likely do not** meet the threshold of significant effects

Norwegian DPA guidance on the AVT Project established that learning analytics used as "decision support" may avoid Article 22 if teachers don't "blindly follow" system results, procedures demonstrate genuine human oversight, and users receive training to consider recommendations in context. However, the Italian DPA's **€200,000 fine against Bocconi University** for remote proctoring demonstrates enforcement when students are not properly informed of automated processing.

Data minimization creates fundamental tension with knowledge tracing's data-intensive requirements. Essend Group LLC CNIL recommendations for AI development emphasize defining objectives before data collection, determining minimum data required for model performance, and considering data-efficient techniques including federated learning, transfer learning, synthetic data generation, and differential privacy.

DPIAs are mandatory for most adaptive learning systems. Article 35 triggers include evaluation/scoring, automated decisions with significant effects, systematic monitoring, sensitive data processing, large-scale processing, vulnerable data subjects (children), and innovative technology (AI/ML)—adaptive learning typically satisfies multiple triggers. CNIL CNIL's position confirms that for high-risk AI systems under the EU AI Act involving personal data, DPIA is "in principle necessary." CNIL

Unified architectures must satisfy the stricter GDPR standard to achieve dual compliance

Multi-tenant EdTech platforms operating across jurisdictions should implement a layered privacy architecture addressing both frameworks simultaneously. The recommended reference architecture includes:

Ingestion Layer: Consent management service handling GDPR Articles 6-7 and COPPA, jurisdiction detection and routing, and data classification distinguishing PII from education records. **Data Governance Layer**: Role-based access control aligned with FERPA's school official model, purpose limitation enforcement per GDPR Article 5, encryption (AES-256 at rest, TLS 1.3 in transit), and automatic audit logging per EU AI Act requirements. **Privacy-Preserving Analytics Layer**: Federated learning module with on-device training option, differential privacy engine with ε-budget management, pseudonymization service (reversible for data subject rights), and k-anonymity/l-diversity processing.

For anonymization meeting both standards, FERPA requires that for any quasi-identifier combination, at least 4+ other individuals share those same attributes, Harvard Online while GDPR demands data not be re-identifiable "by any means reasonably likely to be used." Harvard Online Research from Harvard and UChicago demonstrates that common de-identification methods remain insufficient against determined adversaries, University of Chicago News with educational data particularly vulnerable due to rich behavioral traces, linkability to external datasets, and longitudinal trajectory matching enabling re-identification.

The EU AI Act classifies education as HIGH-RISK under Annex III, covering AI systems determining access to educational opportunities, evaluating learning outcomes, and monitoring prohibited behavior during assessments. Feedbackfruits arXiv This triggers mandatory requirements including risk management systems, data governance ensuring representative training data, technical documentation, automatic event logging, transparency provisions, **human oversight design requirements**, accuracy and robustness standards, quality management systems, Digitaleducationcouncil and registration in the EU database—with compliance deadlines approaching in 2026-2027.

Algorithmic fairness auditing must address both data-related biases (historical, representation, measurement) and algorithmic biases (learning, mapping, confirmation). arxiv Key fairness metrics for educational AI include statistical parity for equal predictions across demographic groups, equal opportunity for identifying at-risk

students needing intervention, equalized odds for fair prediction of both success and failure, and ABROCA (Absolute Between-ROC Area) for overall classification disparity. Protected characteristics extend beyond traditional categories to include disability status, socioeconomic status, English language learner status, and prior educational achievement that may proxy for other protected characteristics.

Industry standards and enforcement establish practical compliance benchmarks

The **Student Data Privacy Consortium's National Data Privacy Agreement v2** (April 2024) represents the most widely adopted compliance framework, with over 130,000 signed DPAs covering 12,000+ schools and 6,674 education application providers across 28 state alliances. The agreement addresses FERPA, PPRA, and COPPA compliance with standardized expectations, state-specific legislative requirements, audit provisions, and breach notification requirements.

1EdTech's TrustEd Apps certification provides the primary interoperability-focused privacy standard, with the Data Privacy Rubric covering data collected, third-party sharing, security practices, and advertising policies. Learning Tools Interoperability (LTI) Advantage 1.3+ certification requires options to enable/disable personal data sharing, with over 250 certified products including Blackboard, Canvas, Moodle, and Sakai.

Recent enforcement actions establish precedent and compliance expectations. The **Illuminate Education settlement** (November 2025) totaling $5.1 million across California, New York, and Connecticut marked the first enforcement under California's KOPIPA, Connecticut's Student Data Privacy Law, and the second major action under NY Education Law § 2-d— CA The Cyber Express demonstrating state attorneys general's willingness to pursue EdTech companies for failed security measures and false privacy policy claims. DataBreaches.Net The **FTC's Edmodo settlement** (2023) established first-of-kind provisions requiring deletion of algorithms developed using improperly collected children's data. Securiti

The Student Privacy Pledge retired in April 2025 after 10 years and nearly 500 signatories, as state laws now supersede its voluntary commitments. EdSurge The **Common Sense Privacy Seal** launched simultaneously as the new third-party validation program, with 2025 recipients including Google Gemini for Education, Quill.org, Book Creator, Padlet, and Seesaw. Commonsenseprivacy

The **CISA K-12 Secure by Design Pledge** (September 2023) provides the most actionable technical commitments, with signatories including PowerSchool, ClassLink, Clever, and Instructure committing to CISA Security Magazine SSO at no additional charge, free security audit logs, published vulnerability disclosure policies, and named executive responsibility for security. THE Journal

For implementation prioritization, EdTech companies should pursue certifications in order: CISA Secure by Design Pledge (free, publicly measurable), SDPC National Data Privacy Agreement signatory status, Common Sense Privacy Seal evaluation, 1EdTech TrustEd Apps certification, and ISO/IEC 27001/27701 certification for comprehensive security and privacy management systems.

Conclusion: Privacy-preserving educational AI requires integrated technical and compliance approaches

The maturation of privacy-preserving machine learning techniques now enables practical implementation of adaptive learning systems that protect student privacy while maintaining educational utility. **Federated learning with differential privacy** represents the recommended architecture for knowledge tracing and recommendation systems, (Google Research) achieving within 3 percentage points of centralized accuracy while keeping sensitive behavioral data distributed.

Key implementation guidance:

- For knowledge tracing, implement FDKT framework with FedProx aggregation on ASSISTments-style data, targeting $\varepsilon = 6\text{-}7$ for approximately 1% AUC degradation
- For classroom analytics with small cohorts, apply temporal aggregation and hierarchical reporting rather than high-noise per-query mechanisms
- For dual FERPA-GDPR compliance, design to GDPR's stricter standards (explicit consent, DPIA, Article 22 safeguards) which generally satisfy FERPA requirements
- For EU AI Act preparation, classify educational AI applications by risk level and implement mandatory human oversight, documentation, and fairness auditing before 2026-2027 deadlines

The regulatory landscape continues evolving rapidly, with state-level enforcement in the US increasingly filling FERPA's gaps and EU AI Act requirements adding substantial obligations for high-risk educational AI. EdTech platforms that proactively implement privacy-by-design architectures, formal fairness auditing, and comprehensive documentation position themselves for both regulatory compliance and competitive advantage in an increasingly privacy-conscious market.