Client-Side Privacy Filtering: A Technical Architecture for GDPR-Compliant AI in Education

Authors:

- Noman Shah, MSc, FBCS, PMP (CEO, XEROTECH LTD; Visiting Faculty, Artificial Intelligence University)
- Adnan Ali Khan (Research Associate, Artificial Intelligence University)

Affiliation: Artificial Intelligence University, London, United Kingdom

Correspondence: noman@xerotech.io

Publication Date: December 2025

DOI: 10.xxxx/aiu.2025.csf.001 [Pending Assignment]

Keywords: Client-side filtering, PII redaction, GDPR compliance, AI in education, privacy-preserving AI, edge computing, data protection, UK Data Protection Act 2018

Word Count: 4,500

Abstract

The adoption of artificial intelligence in educational institutions faces a critical barrier: the inability to reconcile AI capability with data protection requirements. Current solutions rely on third-party trust, including contracts, policies, and enterprise agreements, that still require sensitive data to leave institutional control. This paper presents a client-side privacy filtering architecture that detects and redacts personally identifiable information (PII) before data transmission, ensuring near air-gapped privacy while maintaining full AI functionality. Implemented in the CallGPT platform, this approach addresses UK GDPR and Data Protection Act 2018 requirements through browser-level filtering using pattern-based detection for UK-specific identifiers including National Insurance numbers, NHS numbers, and postcodes. Our findings demonstrate that client-side filtering provides a technically viable solution to the privacy-AI trade-off that has constrained institutional AI adoption across the education sector.

1. Introduction

Artificial intelligence has emerged as a transformative technology in education, offering potential benefits in personalised learning, administrative efficiency, and accessibility support (Department for Education, 2025). However, educational institutions face a fundamental dilemma: the same AI tools that promise productivity gains require access to data that institutions are legally and ethically obligated to protect.

The scale of this challenge is significant. According to Pew Research Center (2024), 81% of consumers believe information collected by AI companies will be used in ways they are uncomfortable with. In the education sector specifically, the UK Department for Education's June 2025 guidance explicitly recommends that "schools should not use personal data in any AI tools" (GOV.UK, 2025). Yet this recommendation, while prudent, effectively prohibits the use of AI for any task involving student or staff information.

The result is what we term the "privacy-AI paralysis": institutions either avoid AI entirely, falling behind in capability, or adopt it with inadequate safeguards, exposing themselves to regulatory risk. A 2023 study by RM Education and Trend Micro found that 52% of UK schools and colleges admitted to not being fully GDPR compliant, suggesting that many institutions are navigating this tension without adequate technical solutions.

This paper presents an alternative approach: client-side privacy filtering that processes data before it leaves the user's browser, eliminating the need to transmit identifiable information to AI providers. We describe the technical architecture, evaluate its effectiveness against UK-specific data protection requirements, and discuss implications for AI adoption in education.

1.1 Research Objectives

This paper aims to:

1. Analyse the current barriers to AI adoption in UK educational institutions with specific focus on data protection requirements
2. Present a technical architecture for client-side PII detection and redaction
3. Evaluate the effectiveness of pattern-based filtering for UK-specific identifiers
4. Discuss the implications for GDPR and UK Data Protection Act 2018 compliance
5. Propose a framework for privacy-preserving AI deployment in education

---

2. Literature Review

2.1 The Privacy-AI Tension in Education

The integration of AI in education has accelerated significantly since the public release of generative AI tools in late 2022. The UK government's AI Opportunities Action Plan identifies education as an area where AI could have positive impact, and the Department for Education has invested £4 million in developing AI tools for teachers (DfE, 2024).

However, this enthusiasm is tempered by substantial privacy concerns. An Ofsted investigation into early adopter schools (Spring 2025) found that "generative AI systems collect and analyse large data sets and there is significant potential for breaches of data privacy" (Ofsted, 2025). The investigation emphasised that robust safeguarding and governance frameworks are essential prerequisites for AI adoption.

The KPMG Global Survey (2024) found that 63% of consumers are concerned about the potential for generative AI to compromise privacy through data breaches or unauthorised access. More specifically, Cisco's 2024 Benchmark Study revealed that 91% of organisations acknowledge they need to do more to reassure customers about how their data is used with generative AI.

2.2 Current Approaches to AI Privacy

Existing approaches to managing AI privacy in institutional settings typically fall into three categories:

Enterprise AI Solutions: Platforms such as ChatGPT Enterprise and Azure OpenAI offer contractual guarantees that user data will not be used for model training. While these solutions address some concerns, they still require data transmission to third-party servers, relying on trust-based compliance rather than technical enforcement.

Data Processing Agreements: Under GDPR Article 28, organisations can establish data processing agreements with AI providers specifying data handling requirements. However, enforcement relies on contractual mechanisms rather than technical controls, and organisations retain liability for any breaches.

Usage Policies and Training: Many institutions implement policies prohibiting the input of sensitive data into AI tools. Research indicates this approach has limited effectiveness. Protecto (2025) found that approximately 15% of employees have pasted sensitive information including PII and financial data into public large language models despite organisational policies.

2.3 Edge Computing and Privacy-Preserving AI

The limitations of trust-based approaches have driven interest in technical solutions that enforce privacy at the point of data processing. Edge computing, which involves processing data closer to the source rather than in centralised cloud infrastructure, offers a paradigm for privacy preservation.

Research from Tokyo University of Science (2024) demonstrated that edge devices can achieve up to 90.2% accuracy in complex tasks while maintaining complete data privacy through local processing. The edge AI market is projected to grow at 33.9% CAGR between 2024 and 2030, driven significantly by demand for privacy-preserving processing (Market Research, 2024).

Client-side filtering represents an extension of edge computing principles to browser-based applications. By processing data within the user's browser before any network transmission, client-side approaches can provide mathematical guarantees that sensitive data never reaches external servers.

2.4 Regulatory Framework

Educational institutions in the UK operate under multiple overlapping data protection requirements:

UK GDPR and Data Protection Act 2018: Requires lawful basis for processing personal data, implementation of appropriate technical and organisational measures, and data minimisation principles. Article 25 specifically mandates "data protection by design and by default."

Information Commissioner's Office Guidance: The ICO's guidance on AI and data protection (updated March 2023) emphasises the importance of ensuring AI tools comply with data protection regulations, with particular attention to children's data under the Age Appropriate Design Code.

Department for Education Guidance: The DfE's generative AI guidance (June 2025) requires schools to conduct Data Protection Impact Assessments (DPIAs) for any AI tool processing pupil data and recommends avoiding personal data in AI tools entirely where possible.

Keeping Children Safe in Education (KCSIE) 2025: For the first time explicitly addresses AI, requiring schools to adopt appropriate filtering and monitoring of AI tools as part of broader safeguarding policies.

The regulatory environment creates clear demand for technical solutions that can demonstrate compliance by design rather than relying solely on policies and procedures.

2.5 GDPR Enforcement in Education

The consequences of non-compliance are substantial and increasing. According to the CMS GDPR Enforcement Tracker Report (2025), data protection authorities across 25 countries have imposed 270 fines on public and educational institutions totalling over €29.3 million.

Notable cases include:

- A Norwegian K-12 school district fined $300,589 for a data breach stemming from inadequately secured third-party software (2023)
- An Italian university fined €200,000 for remote proctoring software that failed to adequately inform students about data processing and transferred data to US servers without appropriate safeguards (2022)
- The University of Greenwich (UK) fined £120,000 for a security breach affecting nearly 20,000 students, staff, and alumni (2018)

These enforcement actions demonstrate that educational institutions face real financial and reputational consequences for data protection failures, and that third-party software is frequently the source of compliance breaches.

---

3. The Privacy Transmission Problem

3.1 Defining the Problem

We identify a fundamental limitation in current AI privacy approaches that we term the "Privacy Transmission Problem": all existing solutions assume that data will, at some point, leave the institution's direct control.

Consider the data flow in a typical AI interaction:

```
User Input → Network Transmission → AI Provider Server → Model Processing → Response
```

At each stage after user input, the institution loses direct control over the data. Even with enterprise agreements, encryption in transit, and contractual guarantees, the data physically exists on third-party infrastructure and is subject to:

- Potential security breaches at the provider
- Legal jurisdiction of the provider's host country
- Changes in provider policies or ownership
- Subpoena or government access requests
- Accidental retention or logging

For educational institutions handling children's data, health information, and other sensitive categories, this loss of control creates irreducible risk regardless of contractual protections.

3.2 The Trust-Based Model's Limitations

Current enterprise AI solutions operate on a trust-based model:

1. Institution trusts provider's contractual commitments
2. Institution trusts provider's security implementation
3. Institution trusts provider's compliance with data retention policies
4. Institution trusts provider's jurisdiction will protect data appropriately

This model has several weaknesses:

Verification Difficulty: Institutions cannot independently verify that providers are honouring commitments. Audit rights exist in theory but are rarely exercised in practice.

Liability Retention: Under GDPR, the data controller (the institution) retains liability even when using compliant processors. Third-party failures become institutional failures.

Dynamic Risk: Provider policies, security postures, and ownership can change. A compliant provider today may not remain compliant tomorrow.

Scope Limitations: Enterprise agreements typically cover intentional data processing but may not address incidental data exposure through logs, debugging systems, or infrastructure components.

3.3 Requirements for a Technical Solution

An effective technical solution to the Privacy Transmission Problem must:

1. **Prevent transmission** of identifiable data, not merely secure it during transmission
2. **Operate at the source** before data enters any network pathway
3. **Maintain functionality** so users can still accomplish AI-assisted tasks
4. **Preserve context** so the AI receives sufficient information to provide useful responses
5. **Provide auditability** so institutions can demonstrate compliance
6. **Support configurability** as different data types may require different handling based on institutional policy

4. Proposed Solution: Client-Side Privacy Filtering

4.1 Architecture Overview

We propose a client-side privacy filtering architecture that intercepts user input within the browser, detects personally identifiable information using pattern matching, and redacts sensitive data before any network transmission occurs.

```
User Input → Client-Side Filter → Sanitised Input → Network → AI Provider
    ↓           ↓
Original stored    PII detected
(local display)    and redacted
```

The key distinction from existing approaches is that filtering occurs entirely within the browser environment. The AI provider never receives identifiable data because it is removed before transmission, not because contracts prohibit its use after receipt.

4.2 Technical Implementation

The client-side filtering system operates through the following components:

Detection Engine: A pattern-matching system using regular expressions optimised for UK-specific identifiers. The engine scans user input in real-time, identifying data patterns that match known PII formats.

Redaction Processor: When PII is detected, the processor replaces identifiable data with standardised placeholders (e.g., `[NI_REDACTED]`, `[NHS_REDACTED]`). These placeholders preserve semantic meaning for the AI while removing identifying information.

Dual Storage System: The original user input is stored locally for display purposes, while the sanitised version is transmitted to AI providers. This ensures users can see their original text while guaranteeing that only redacted content reaches external systems.

Audit Logger: All redaction events are logged with metadata (timestamp, redaction type, user identifier) but without the actual redacted values. This provides an audit trail for compliance demonstration without creating a secondary privacy risk.

4.3 Data Classification Tiers

The system implements a tiered approach to data classification, recognising that different data types carry different risk profiles:

Tier 1: High Sensitivity (Always Redacted)

- UK National Insurance numbers (format: XX 00 00 00 X)
- Credit and debit card numbers (16-digit patterns with common separators)
- UK Passport numbers (9-digit format)
- UK Driving licence numbers (16-character alphanumeric)

Tier 2: Medium Sensitivity (Redacted by Default)

- UK mobile and landline phone numbers (+44, 07xxx, 01xxx, 02xxx formats)
- UK postcodes (standard format patterns)
- Email addresses (standard email pattern matching)

Tier 3: Contextual Sensitivity (Configurable)

- Bank account numbers (8-digit patterns with higher false positive risk)
- Sort codes (6-digit patterns that may match dates)
- NHS numbers (10-digit patterns that may match phone numbers)

Tier 3 items are disabled by default due to false positive risk but can be enabled by institutions with specific requirements.

4.4 Pattern Specifications

The detection engine employs the following regular expression patterns for UK-specific identifiers:

National Insurance Number:

```regex
/[A-CEGHJ-PR-TW-Z]{2}\s?\d{2}\s?\d{2}\s?\d{2}\s?[A-D]/gi
```

This pattern matches the official HMRC format: two prefix letters (excluding certain combinations), six digits in three pairs, and a suffix letter A-D.

UK Phone Numbers:

```regex
/(\+44\s?|0)7\d{3}\s?\d{6}/g  // Mobile
/(\+44\s?|0)[12]\d{2,3}\s?\d{3}\s?\d{4}/g  // Landline
```

UK Postcodes:

```regex
/\b[A-Z]{1,2}\d[A-Z\d]?\s*\d[A-Z]{2}\b/gi
```

Email Addresses:

```regex
/[a-zA-Z0-9._%+-]+@[a-zA-Z0-9.-]+\.[a-zA-Z]{2,}/g
```

4.5 User Context Exclusion

A critical design consideration is that users should be able to reference their own information naturally. The system implements user context exclusion: the authenticated user's own email address and name are excluded from redaction, allowing natural self-reference while still protecting third-party data.

```
User: "My email is john@school.edu" → Transmitted as-is (user's own data)
User: "Contact jane@parent.com" → Transmitted as "Contact [EMAIL_REDACTED]"
```

This approach balances privacy protection with usability, recognising that users have legitimate reasons to reference their own information.

4.6 Implementation in CallGPT

The client-side filtering architecture has been implemented in CallGPT, a multi-provider AI platform. The implementation includes:

- Real-time filtering with sub-100ms processing latency
- Support for all Tier 1 and Tier 2 identifiers
- Configurable Tier 3 detection
- User notification when redaction occurs (privacy shield indicator)
- Audit logging compliant with GDPR Article 30 record-keeping requirements
- Dual storage preserving original content for user reference

The system processes all user messages before API calls to any AI provider (OpenAI, Anthropic, Google, etc.), ensuring consistent privacy protection regardless of which model is selected.

---

5. Evaluation

5.1 Detection Accuracy

We evaluated the detection system against a test corpus of 10,000 synthetic messages containing UK PII in various formats. Results for Tier 1 and Tier 2 identifiers:

Identifier Type	Precision	Recall	F1 Score
National Insurance	99.2%	97.8%	98.5%
Credit Card	98.7%	99.1%	98.9%
UK Mobile Phone	97.3%	96.5%	96.9%
UK Postcode	95.8%	94.2%	95.0%
Email Address	99.4%	98.9%	99.1%

Lower precision for postcodes reflects the challenge of distinguishing postcodes from similar alphanumeric patterns in technical content. This is an acceptable trade-off given the relatively low sensitivity of postcode data compared to other identifiers.

5.2 Performance Impact

Client-side processing adds minimal latency to user interactions:

Message Length	Processing Time	Overhead vs. Unfiltered
Short (<100 chars)	12ms	+8ms
Medium (100-500 chars)	24ms	+15ms
Long (500-2000 chars)	47ms	+28ms

These performance characteristics are well within acceptable bounds for interactive use, with processing completing before typical network latency.

5.3 Functional Preservation

A key concern with any redaction system is whether AI responses remain useful after PII removal. We conducted qualitative evaluation across common educational use cases:

Administrative Tasks: Tasks such as drafting parent communications, generating report card comments, and creating student feedback remained fully functional with redacted data. The AI could generate appropriate content using placeholders, which users could then populate with actual names.

Data Analysis: Aggregate analysis tasks (e.g., "analyse these student performance patterns") functioned normally as statistical patterns are preserved even with individual identifiers redacted.

Personalised Content: Tasks requiring specific individual reference (e.g., "write a letter to John Smith about his progress") required users to add names to AI-generated template content, adding a manual step but maintaining privacy.

5.4 Compliance Assessment

The client-side filtering architecture addresses key GDPR requirements:

Article 5, Data Minimisation: By redacting PII before transmission, the system ensures that AI providers receive only the minimum data necessary to provide the service.

Article 25, Data Protection by Design: Privacy protection is implemented at the technical level, not relying solely on policies or procedures.

Article 32, Security of Processing: By preventing PII transmission, the system eliminates entire categories of security risk associated with third-party data processing.

Articles 44-49, International Transfers: Redacted data does not constitute personal data under GDPR, potentially simplifying international transfer considerations when using US-based AI providers.

---

6. Discussion

6.1 Advantages of Client-Side Filtering

The client-side approach offers several advantages over existing privacy solutions:

Technical Enforcement: Privacy is enforced through code, not contracts. Institutions do not need to trust third-party commitments because sensitive data never leaves their control.

Reduced Liability: By preventing PII transmission, institutions reduce their exposure under GDPR processor liability provisions. The AI provider never becomes a data processor for the redacted information.

Audit Simplicity: Demonstrating compliance becomes straightforward. Institutions can show that their technical architecture prevents PII transmission rather than relying on third-party audit reports.

Provider Agnosticism: The filtering layer operates independently of the AI provider, allowing institutions to switch providers or use multiple providers without renegotiating data processing agreements for each.

6.2 Limitations

The approach has several limitations that institutions should consider:

Pattern-Based Detection: The system relies on pattern matching, which cannot detect PII that doesn't match expected formats or PII described in prose (e.g., "my national insurance number is alpha bravo one two..."). Future development may incorporate natural language processing for contextual detection.

UK-Specific Optimisation: Current patterns are optimised for UK identifiers. Institutions with international students or staff may require additional patterns for other jurisdictions.

User Circumvention: Determined users could potentially format data to avoid detection. The system is designed to prevent accidental exposure, not malicious circumvention.

Functional Constraints: Some AI use cases genuinely require identified data (e.g., generating personalised certificates with correct names). These cases require alternative workflows or manual data insertion after AI generation.

6.3 Implications for Educational AI Adoption

The availability of technically-enforced privacy solutions has significant implications for AI adoption in education:

Enabling Cautious Adoption: Institutions that have avoided AI due to privacy concerns may find client-side filtering provides sufficient assurance to proceed with controlled implementation.

Simplifying Governance: Privacy-by-design approaches reduce the governance burden, as institutions need not maintain complex data processing agreements or conduct ongoing provider audits.

Supporting Regulatory Compliance: As regulators increasingly focus on AI in education (evidenced by KCSIE 2025 updates and DfE guidance), technical privacy controls provide defensible compliance positions.

Levelling Access: Smaller institutions without resources for complex enterprise AI agreements can achieve equivalent privacy protection through technical means.

6.4 Future Research Directions

Several areas warrant further investigation:

Machine Learning-Based Detection: Incorporating ML models for contextual PII detection could address limitations of pattern-based approaches, though this introduces complexity around model training data and false positive management.

Federated Learning Integration: Combining client-side filtering with federated learning approaches could enable model improvement without centralised data collection.

Cross-Jurisdictional Patterns: Developing comprehensive pattern libraries for EU, US, and other jurisdictions would extend applicability beyond UK-specific use cases.

Longitudinal Compliance Studies: Tracking regulatory interpretations of client-side filtering approaches as AI-specific regulations (such as the EU AI Act) mature.

---

7. Conclusion

The tension between AI capability and data protection has created significant barriers to AI adoption in educational institutions. Current approaches relying on enterprise agreements, data processing contracts, and usage policies address symptoms rather than root causes. They manage what happens after sensitive data leaves institutional control rather than preventing that departure.

Client-side privacy filtering offers a fundamentally different approach: technical enforcement of privacy at the point of data origin. By detecting and redacting personally identifiable information within the browser before any network transmission, this architecture eliminates entire categories of privacy risk while preserving AI functionality for the majority of educational use cases.

Our implementation in the CallGPT platform demonstrates the technical viability of this approach, achieving high detection accuracy for UK-specific identifiers with minimal performance overhead. The architecture aligns with GDPR principles of data minimisation and protection by design, potentially simplifying compliance for educational institutions seeking to adopt AI tools.

As the regulatory landscape continues to evolve with the EU AI Act, updated UK guidance, and increasing enforcement activity, technically-enforced privacy solutions offer educational institutions a defensible path forward. The question for institutions is no longer whether to adopt AI, but how to do so in ways that genuinely protect the students and staff whose data they are obligated to safeguard.

Client-side filtering does not solve every privacy challenge in educational AI, but it addresses the most fundamental one: ensuring that the choice to protect student data is enforced by architecture, not merely promised by policy.

References

1. Cisco. (2024). *2024 Data Privacy Benchmark Study*. Cisco Systems. Retrieved from https://www.cisco.com/c/en/us/about/trust-center/data-privacy-benchmark-study.html
2. CMS Law. (2025). *GDPR Enforcement Tracker Report 2024/2025: Public Sector and Education*. Retrieved from https://cms.law/en/int/publication/gdpr-enforcement-tracker-report/public-sector-and-education
3. Department for Education. (2025). *Generative artificial intelligence (AI) in education*. GOV.UK. Retrieved from https://www.gov.uk/government/publications/generative-artificial-intelligence-in-education
4. Department for Education. (2025). *Data protection in schools: Generative artificial intelligence (AI) and data protection in schools*. GOV.UK. Retrieved from https://www.gov.uk/guidance/data-protection-in-schools/generative-artificial-intelligence-ai-and-data-protection-in-schools
5. DLA Piper. (2024). *GDPR Fines and Data Breach Survey: January 2024*. Retrieved from https://www.dlapiper.com/en/insights/publications/2024/01/dla-piper-gdpr-fines-and-data-breach-survey-january-2024
6. Endpoint Protector. (2023). *Educational Institutions: How to Ensure Data Compliance and Security*. Retrieved from https://www.endpointprotector.com/blog/educational-institutions-how-to-ensure-data-compliance-and-security/
7. GDPR Sentry. (2025). *Education and UK GDPR in 2025: Seven Years On, Are We Getting It Right?* Retrieved from https://gdprsentry.com/education-and-uk-gdpr-in-2025-seven-years-on-are-we-getting-it-right/
8. Information Commissioner's Office. (2023). *Guidance on AI and Data Protection*. ICO. Retrieved from https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/guidance-on-ai-and-data-protection/
9. KPMG. (2024). *Generative AI and the enterprise: Global insights on trust and adoption*. KPMG International. Retrieved from https://kpmg.com/xx/en/home/insights/2024/01/generative-ai-survey.html
10. Kumar, L. (2023). *Edge Computing for AI and ML: Enhancing Performance and Privacy in Data Analysis*. SSRN Electronic Journal. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5052105
11. Muckle LLP. (2025). *Your guide to using generative artificial intelligence in education*. Retrieved from https://www.muckle-llp.com/insights/legal-commentary/gen-ai-in-education/
12. Nuffield Foundation & Ada Lovelace Institute. (2025). *A learning curve? The role of AI in education: Opportunities and challenges*. Retrieved from https://www.nuffieldfoundation.org/news/the-role-of-ai-in-education-opportunities-and-challenges
13. Ofsted. (2025). *'The biggest risk is doing nothing': insights from early adopters of artificial intelligence in schools and further education colleges*. GOV.UK. Retrieved from https://www.gov.uk/government/publications/ai-in-schools-and-further-education-findings-from-early-adopters
14. Pearson. (2025). *The Pearson School Report 2025: UK Educators Voice Concerns Over AI Readiness*. Retrieved from https://www.pearson.com/en-gb/schools/insights-and-events/topics/school-report.html

15. Pew Research Center. (2024). *Public views on AI, privacy, and data use*. Retrieved from https://www.pewresearch.org/internet/2024/03/public-awareness-and-attitudes-about-ai/

16. Protecto. (2025). *AI Data Privacy Statistics & Trends 2025*. Retrieved from https://www.protecto.ai/blog/ai-data-privacy-statistics-trends/

17. Relyance AI. (2024). *Consumer AI Trust Survey 2025*. Retrieved from https://www.relyance.ai/consumer-ai-trust-survey-2025

18. SchoolPro TLC. (2025). *Guidance on the use of Generative AI in MATs and Schools*. Retrieved from https://schoolpro.uk/2025/07/guidance-on-the-use-of-generative-ai-in-mats-and-schools/

19. Secureframe. (2025). *110+ Data Privacy Statistics: The Facts You Need To Know In 2025*. Retrieved from https://secureframe.com/blog/data-privacy-statistics

20. Stanford University. (2025). *AI Index Report 2025*. Stanford Institute for Human-Centered Artificial Intelligence. Retrieved from https://aiindex.stanford.edu/report/

21. Structural Learning. (2025). *Creating an AI Policy for Schools: A Practical Guide for 2025*. Retrieved from https://www.structural-learning.com/post/creating-ai-policy-schools-2025

22. Termly. (2025). *54 Revealing AI Data Privacy Statistics*. Retrieved from https://termly.io/resources/articles/ai-statistics/

23. Tony Blair Institute for Global Change. (2025). *Generation Ready: Building the Foundations for AI-Proficient Education in England's Schools*. Retrieved from https://institute.global/insights/public-services/generation-ready-building-the-foundations-for-ai-proficient-education-in-englands-schools

24. UK Government. (2025). *Keeping Children Safe in Education 2025*. Department for Education. Retrieved from https://www.gov.uk/government/publications/keeping-children-safe-in-education--2

25. Usercentrics. (2025). *Over 150 data privacy statistics companies need to know about in 2025*. Retrieved from https://usercentrics.com/guides/data-privacy/data-privacy-statistics/

26. XL Law Consulting. (2023). *GDPR Enforcement Actions: Lessons Learned for Colleges and Universities*. Retrieved from https://www.xllawconsulting.com/post/gdpr-enforcement-actions-against-educational-institutions

---

Bibliography

Primary Legislation and Regulatory Guidance

- Data Protection Act 2018. (2018). UK Public General Acts. https://www.legislation.gov.uk/ukpga/2018/12/contents/enacted
- European Parliament. (2016). Regulation (EU) 2016/679 (General Data Protection Regulation). Official Journal of the European Union. https://eur-lex.europa.eu/eli/reg/2016/679/oj
- European Parliament. (2024). Regulation (EU) 2024/1689 (Artificial Intelligence Act). Official Journal of the European Union. https://eur-lex.europa.eu/eli/reg/2024/1689/oj
- Information Commissioner's Office. (2024). *Children's code: a guide for education settings*. https://ico.org.uk/for-organisations/childrens-code-hub/
- UK Government. (2025). Data (Use and Access) Act 2025. https://www.legislation.gov.uk/ukpga/2025/

Academic Literature

- Mothukuri, V., Parizi, R. M., Pouriyeh, S., Huang, Y., Dehghantanha, A., & Srivastava, G. (2021). A survey on security and privacy of federated learning. *Future Generation Computer Systems*, 115, 619-640. https://doi.org/10.1016/j.future.2020.10.007
- Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology*, 10(2), 1-19. https://doi.org/10.1145/3298981
- Zhang, C., Xie, Y., Bai, H., Yu, B., Li, W., & Gao, Y. (2021). A survey on federated learning. *Knowledge-Based Systems*, 216, 106775. https://doi.org/10.1016/j.knosys.2021.106775
- Zhou, J., Chen, F., & Holzinger, A. (2022). AI ethics and privacy in edge intelligence. *IEEE Transactions on Artificial Intelligence*, 3(5), 734-745. https://doi.org/10.1109/TAI.2022.3145154

Technical Standards

- IAPP. (2024). *Consumer Perspectives of Privacy and Artificial Intelligence*. International Association of Privacy Professionals. https://iapp.org/resources/article/consumer-perspectives-of-privacy-and-ai/
- NIST. (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. National Institute of Standards and Technology. https://www.nist.gov/itl/ai-risk-management-framework

Industry Reports

- Cisco. (2025). *2025 Data Privacy Benchmark Study*. https://newsroom.cisco.com/c/r/newsroom/en/us/a/y2025/m04/cisco-2025-data-privacy-benchmark-study-privacy-landscape-grows-increasingly-complex-in-the-age-of-ai.html
- IBM. (2024). *Cost of a Data Breach Report 2024*. IBM Security. https://www.ibm.com/reports/data-breach
- McKinsey & Company. (2024). *The state of AI in 2024: Generative AI's breakout year*. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai

Appendix A: Implementation Specifications

A.1 System Requirements

- Modern web browser with JavaScript ES6+ support
- Minimum 4GB RAM for client-side processing
- Network connectivity to AI provider APIs

A.2 API Integration

The filtering system integrates with AI provider APIs through a middleware layer:

```javascript
const privacyResult = stripPrivacyIdentifiers(userMessage, {
  enabled: true,
  stripTier1: true,
  stripTier2: true,
  stripTier3: false,
  userContext: {
    email: authenticatedUser.email,
    name: authenticatedUser.fullName
  }
});

// Original stored for display
message.content = userMessage;

// Sanitised version sent to AI
message.contentForAI = privacyResult.sanitised;
```

A.3 Audit Log Format

```json
{
  "timestamp": "2025-12-17T10:30:00Z",
  "userId": "hashed_identifier",
  "sessionId": "session_uuid",
  "redactionCount": 3,
  "redactionTypes": ["NI_NUMBER", "PHONE", "EMAIL"],
  "aiProvider": "anthropic",
  "model": "claude-3-sonnet"
}
```

Note: Actual redacted values are never logged.

Appendix B: Regulatory Mapping

GDPR Article	Requirement	Client-Side Filtering Implementation
Art. 5(1)(c)	Data minimisation	PII removed before transmission
Art. 5(1)(f)	Integrity and confidentiality	Sensitive data never leaves browser
Art. 25(1)	Data protection by design	Privacy enforced architecturally
Art. 25(2)	Data protection by default	Tier 1 & 2 filtering enabled by default
Art. 30	Records of processing	Audit logging without PII storage
Art. 32	Security of processing	Eliminates third-party transmission risk
Art. 35	Data protection impact assessment	Simplified DPIA with reduced processing scope

---

Corresponding Author: Noman Shah, MSc, FBCS, PMP Email: noman@xerotech.io XEROTECH LTD, London, United Kingdom

Author Contributions: N.S. conceived the research, designed the technical architecture, and wrote the manuscript. A.A.K. contributed to the literature review and data analysis. Both authors reviewed and approved the final manuscript.

Conflicts of Interest: Noman Shah is CEO of XEROTECH LTD, which develops the CallGPT platform described in this paper. The research was conducted independently, and technical claims are based on documented system specifications. Adnan Ali Khan declares no conflicts of interest.

Funding: This research received no external funding.

Data Availability: Technical documentation for the CallGPT privacy filtering system is available at https://callgpt.co.uk/privacy-documentation. The test corpus used for evaluation is available from the corresponding author upon reasonable request.

Acknowledgements: The authors thank the Artificial Intelligence University for institutional support and the early adopter institutions who provided feedback during system development.

Citation: Shah, N., & Khan, A. A. (2025). Client-Side Privacy Filtering: A Technical Architecture for GDPR-Compliant AI in Education. *Artificial Intelligence University Research Publications*. https://aiu.ac/research/client-side-privacy-filtering

---