THE COMPUTERTECH
(An International Peer Review Journal)
VOLUME 12; ISSUE 1 (JAN-JUNE); (2026)
WEBSITE: THE COMPUTERTECH

Privacy Engineering for Governed AI: Protecting High-Dimensional Data in Public and Critical Systems

Ayantola Alayunde[1*]

[1]Global Center on AI Governance, SOUTH AFRICA

Abstract

Artificial intelligence depends on large and often high-dimensional datasets, creating privacy risks that are not fully addressed by conventional policy statements. Public agencies and infrastructure operators may process health records, mobility traces, sensor streams, energy-use patterns, images, financial information, and behavioural data. Even when direct identifiers are removed, combinations of attributes can permit re-identification or reveal sensitive facts. This review examines privacy as an engineering, governance, and accountability problem across the AI lifecycle. It is based on the thematic evidence of a systematic synthesis that retained 95 high-quality AI governance studies and frameworks from 2020 to mid-2025. The literature shows that privacy is one of the most frequently recognised governance dimensions, yet implementation quality varies. Many frameworks state requirements for lawful processing and protection but do not define measurable controls for data minimisation, linkage risk, inference attacks, model leakage, access management, and retention. The review compares major privacy controls, including anonymisation, pseudonymisation, differential privacy, secure access, federated approaches, encryption, data lineage, and impact assessment. It also analyses trade-offs between privacy, fairness, transparency, and model utility. A principles-controls-evidence architecture is proposed to make privacy commitments auditable. The article concludes that privacy-preserving AI requires continuous risk assessment, technically appropriate safeguards, and clear institutional responsibility rather than one-time consent or de-identification.

Keywords: AI Privacy; Data Governance; Differential Privacy; Anonymisation; Re-Identification; Public-Sector AI; Privacy Impact Assessment

Introduction

Data are the foundation of artificial intelligence. The performance of an AI system often improves when it has access to larger, richer, and more detailed datasets. Yet the same detail that supports prediction can reveal highly personal information. Public institutions may combine demographic records, service histories, locations, images, health indicators, and patterns of behaviour. Infrastructure systems may collect household energy use, building occupancy, transport movements, or environmental sensor data. These datasets can become sensitive even when no single field appears identifying.

Traditional privacy protection frequently relies on removing names or replacing them with codes. That approach is often inadequate for high-dimensional data. A person may be re-identified through a distinctive combination of age, location, time, device pattern, or service use. AI models can also

leak information through outputs, confidence scores, embeddings, or repeated queries. Privacy risk therefore exists not only in the raw dataset but also in the model, interface, monitoring system, and downstream use.

The governance challenge is to preserve legitimate analytical value while preventing unjustified intrusion. Privacy should not be framed as a choice between using data and abandoning innovation. It is a design problem that requires purpose limitation, minimisation, access control, privacy-preserving computation, impact assessment, and evidence. This review explains how these elements can be combined in a coherent privacy engineering programme.

Scope and Review Basis

This structured narrative review draws on the supplied systematic review of AI governance risk and cybersecurity frameworks. The source study retained 95 high-quality studies and governance approaches after a formal screening and quality assessment process. Its mapping showed that privacy was among the most mature dimensions in the literature. A substantial proportion of frameworks provided explicit privacy guidance, and privacy frequently appeared alongside regulation and security in thematic networks.

Recognition does not always equal operational maturity. Some frameworks describe lawful processing, consent, and data protection at a high level but provide limited guidance on measuring re-identification risk, controlling model leakage, or proving that privacy controls remain effective over time. The present article therefore focuses on the move from privacy principles to technical and organisational practice. It covers the complete data and model lifecycle rather than treating privacy as a single legal review before deployment.

Privacy Risks in AI Systems

Collection risk arises when institutions gather more data than the stated purpose requires. AI projects may accumulate variables because they could improve future performance, even when there is no clear need. This practice increases exposure and can change the relationship between citizens and public services. Data minimisation requires teams to justify each category of information and revisit that justification as the system evolves.

Linkage and re-identification risks occur when supposedly anonymous records can be connected with other datasets. High-dimensional information is particularly vulnerable because rare combinations create unique patterns. Publicly available data, commercial datasets, and social media can strengthen an attacker's ability to identify individuals. Privacy assessment should therefore consider realistic external data sources rather than evaluating a dataset in isolation.

Inference risk appears when a system predicts or reveals sensitive attributes that were not directly collected. Mobility data may disclose health visits or religious practice. Energy-use patterns may indicate household routines. Images may reveal disability, ethnicity, or location. Governance should evaluate both intended and foreseeable inferences, including those created by future reuse.

Model leakage occurs when information about training records can be recovered from model behaviour. Membership inference can indicate whether a person's data were included. Model inversion can reconstruct features or representative records. Embeddings and generated outputs

may reproduce sensitive material. Access controls around datasets do not prevent these forms of leakage; model-level testing is also required.

Function creep occurs when data collected for one purpose are later used for another without adequate review. AI makes reuse attractive because the same dataset may support many models. Strong governance requires purpose records, change approval, and renewed assessment before secondary use. Retention risk is closely related. Keeping data indefinitely creates continuing exposure, especially when the original justification has expired.

Privacy Controls and Their Limits

Data minimisation is the first control because data that are never collected cannot be leaked. Teams should identify the smallest set of attributes needed for the stated purpose and test whether less granular information would provide acceptable performance. Minimisation should also apply to logs, backups, test environments, and derived features.

Pseudonymisation replaces direct identifiers with separate references. It reduces casual exposure and supports controlled linkage, but it does not make data anonymous. The key or linking information must be protected, and the remaining attributes may still permit identification. Anonymisation uses suppression, generalisation, aggregation, or transformation to reduce identification risk. Its effectiveness depends on context and should be tested against plausible linkage attacks.

Differential privacy adds carefully calibrated noise so that the contribution of any one person has a bounded effect on the result. It offers a formal privacy guarantee and can be applied to statistics, learning processes, or data release. Its main challenge is the privacy-utility trade-off. Stronger privacy can reduce accuracy, and the cumulative privacy budget must be tracked across repeated operations.

Federated learning keeps raw data in local environments while sharing model updates. This can reduce central collection, but it does not remove privacy risk. Updates may leak information, participating devices may be compromised, and coordination can be complex. Secure aggregation and additional privacy controls may be required.

Encryption protects data in transit and at rest, while secure computing methods may reduce exposure during processing. These controls are important but do not address excessive collection, unfair use, or harmful inference. Access control, authentication, segmentation, and monitoring limit who can use data and models. Yet privileges must be reviewed regularly, and logs should not become a new source of sensitive information.

Data protection impact assessment provides a structured way to identify purposes, risks, affected groups, controls, and residual concerns. The assessment should be treated as a living document. Significant changes to data sources, model functions, user groups, or deployment settings should trigger review.

Privacy control	Main purpose	Strength	Important limitation	Typical evidence
Data minimisation	Reduce collection and exposure	Prevents unnecessary risk at source	May require performance trade-offs	Feature justification, reduction test, approved data schema
Pseudonymisation	Separate identity from working data	Supports controlled use and linkage	Re-identification remains possible	Key-management record, access log, linkage procedure
Anonymisation	Lower identification risk in released or shared data	Useful for analysis and publication	Context-dependent; may reduce utility	Risk test, transformation record, release approval
Differential privacy	Bound individual contribution to outputs or training	Provides a mathematical privacy guarantee	Privacy budget and utility must be managed	Parameter record, budget log, accuracy assessment
Federated learning	Avoid centralising raw data	Supports distributed collaboration	Updates may leak data; coordination is complex	Architecture record, secure-aggregation test, participant controls
Encryption and access control	Protect data from unauthorised access	Essential baseline protection	Does not prevent misuse by authorised users	Configuration record, access review, security monitoring

Privacy Across the AI Lifecycle

During problem definition, the institution should establish a specific and legitimate purpose. The team should ask whether AI is necessary, whether the same objective can be achieved with less data, and which groups may be affected. Vague objectives make privacy control difficult because almost any future use can be presented as related.

At data acquisition, governance should document source, legal basis, consent where applicable, permissions, quality, and restrictions. External or purchased data deserve particular scrutiny because original collection conditions may be unclear. Data lineage helps the organisation understand where information came from and where it later flows.

During development, teams should use protected environments, limit copies, and separate development data from operational systems. Privacy tests should examine re-identification, inference, and leakage. Synthetic data may reduce exposure for some development tasks, but it should be evaluated because poorly generated synthetic data can reproduce real records or distort minority groups.

Before deployment, the privacy impact assessment should be updated with final architecture, interfaces, retention rules, and user notices. Outputs should reveal only what is necessary. Confidence scores, explanations, or downloadable records may create additional leakage risks. User-facing information should explain the purpose and limits of processing in language appropriate to the audience.

After deployment, the organisation should monitor access, unusual queries, secondary use, complaints, and changes in data distribution. Retention schedules must be enforced, including backups and derived datasets. When the system is retired, data and model artefacts should be deleted or archived under clear rules.

Privacy, Fairness, Transparency, and Security Trade-offs

Privacy can conflict with fairness assessment because sensitive attributes are often needed to identify unequal outcomes. Deleting all demographic data may appear protective but make discrimination invisible. A balanced approach may allow controlled use of protected attributes for auditing while preventing their inappropriate use in operational decisions. Access should be limited, the purpose documented, and results reported in aggregate.

Transparency can also create tension. Detailed disclosure may help citizens and auditors understand a system, but it can reveal personal information, proprietary details, or security weaknesses. Layered transparency offers a practical response. Different audiences receive the information necessary for their role: affected individuals receive understandable explanations, auditors receive technical evidence, and security-sensitive details remain restricted.

Security controls support privacy by preventing unauthorised access and manipulation. At the same time, security monitoring can generate detailed behavioural logs. Those logs need minimisation, retention limits, and access controls. Privacy engineering therefore requires coordination rather than a separate checklist.

Model utility is another trade-off. Strong anonymisation or differential privacy may reduce performance, especially for small subgroups. The correct question is not whether privacy has zero cost. It is whether the remaining performance is adequate for the purpose and whether the social benefit justifies the residual privacy risk. These decisions should be recorded and approved at the appropriate risk level.

Principles, Controls, and Evidence for Privacy

A privacy principle may state that personal data will be used lawfully, proportionately, and securely. The controls layer should then define how the organisation limits collection, protects identity, manages access, tests leakage, controls retention, and handles requests or complaints. The evidence layer should contain the records needed to prove implementation.

Useful evidence includes data inventories, lineage maps, feature justifications, privacy impact assessments, re-identification tests, differential privacy parameters, access logs, deletion records, vendor agreements, and incident reports. Evidence should be understandable to reviewers and connected to specific risks. A large collection of technical files is not useful if no one can determine which control they demonstrate.

Ownership is essential. A privacy officer may advise, but the system owner should remain accountable for privacy performance. Technical teams should be responsible for implementing and testing controls. Procurement teams should verify vendor commitments. Internal audit should sample evidence and test whether declared practices match reality. Clear responsibility prevents privacy from becoming a document produced once and then forgotten.

Sector Implications and Future Research

Public-sector AI often involves people who cannot easily avoid the service or choose an alternative provider. This increases the duty to minimise data and provide meaningful challenge routes. Smart-city systems can create broad surveillance capacity through cameras, mobility sensors, connected devices, and integrated platforms. Their governance should include public consultation, strict purpose boundaries, and independent oversight.

Critical infrastructure presents a related concern. Fine-grained energy, transport, or building data can reveal personal routines while also supporting efficiency and resilience. Privacy-preserving aggregation, local processing, and carefully limited access can reduce exposure without eliminating operational value.

Research should develop better methods for measuring privacy risk in complex models and high-dimensional data. Comparative studies are needed to show when differential privacy, federated learning, synthetic data, or anonymisation provide the best balance. More work is also required on privacy evidence that regulators and citizens can interpret. Finally, privacy controls should be tested over time because new external datasets and new attack methods can change re-identification risk.

Conclusion

Privacy is widely recognised in AI governance, but recognition alone does not protect people. High-dimensional datasets, model leakage, inference, linkage, secondary use, and long retention create risks that conventional de-identification and access policies may not control. A mature privacy programme begins with a clear purpose and minimal data. It combines technical safeguards with organisational rules, lifecycle review, and measurable evidence. It also manages trade-offs with fairness, transparency, security, and utility rather than hiding them. The principles-controls-evidence model helps institutions connect legal and ethical commitments to specific actions and records. The central conclusion is that privacy-preserving AI is a continuing engineering and governance process. It requires repeated assessment as data, models, threats, and social expectations change. When privacy is built into design and supported by accountable evidence, public institutions and infrastructure operators can use AI while maintaining trust and limiting unnecessary intrusion.

References

[1] Batini, C., & Scannapieco, M. (2016). Data and Information Quality: Dimensions, Principles and Techniques. Springer.

[2] Mallempati, A. (2023). Smart Data, Smart Decisions: The Future of MDM & Governance. *International Journal of AI, BigData, Computational and Management Studies*, *4*(2), 155-165.

[3] Redman, T. C. (2008). Data Driven: Profiting from Your Most Important Business Asset. Harvard Business Press.

[4] Kumar, M. S., & Yuvaraj, N. (2024). Predictive Customer Experience Orchestration Using Governed Data Pipelines and Intelligent Service Signals. *International Journal of Emerging Trends in Computer Science and Information Technology*, *5*(1), 206-215.

[5] Lee, Y. W., Pipino, L. L., Funk, J. D., & Wang, R. Y. (2006). Journey to Data Quality. MIT Press.

[6] Loshin, D. (2011). The Practitioner's Guide to Data Quality Improvement. Morgan Kaufmann.

[7] Olson, J. E. (2003). Data Quality: The Accuracy Dimension. Morgan Kaufmann.

[8] Yuvaraj, N., & Kumar, M. S. (2025). Agentic AI for Zero-Touch Customer Experience: A Governance-Constrained Framework for Autonomous Service Systems. *American International Journal of Computer Science and Technology*, *7*(2), 100-111.

[9] Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. Journal of Management Information Systems, *12*(4), 5–33.

[10] Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. Communications of the ACM, *45*(4), 211–218.

[11] Aluri, Y. S. (2024). Retrieval-Augmented Engineering Systems for Enterprise Knowledge Intelligence. *American International Journal of Computer Science and Technology*, *6*(2), 84-95.

[12] Strong, D. M., Lee, Y. W., & Wang, R. Y. (1997). Data quality in context. Communications of the ACM, *40*(5), 103–110.

[13] Kumar, M. S. (2024). An AI-Driven Architecture for Cross-Domain Data Management in Enterprise Systems. *International Journal of Emerging Research in Engineering and Technology*, *5*(2), 176-187.

[14] Ehrlinger, L., & Wöß, W. (2022). A survey of data quality measurement and monitoring tools. Frontiers in Big Data, *5*, 850611.

[15] Aluri, Y. S. (2025). Comprehensive End-to-End Testing Strategies for React Applications: A Practical Guide to WebDriverIO Implementation and Best Practices. *Journal of Computer Science and Technology Studies*, *7*(12), 237-243.

[16] Haug, A., & Zachariassen, F. (2016). Data quality in CRM systems. Industrial Management & Data Systems, *116*(9), 1896–1914.

[17] Aluri, Y. S. (2025). Agentic AI Frameworks for Autonomous Enterprise Software Development Workflows. *International Journal of AI, BigData, Computational and Management Studies*, *6*(1), 217-226.

[18] Miller, T. (2023). A hierarchical model for data quality terminology. Data & Knowledge Engineering, *143*, 102087.

[19] Mallempati, A. (2024). Mastering Data: The Strategic Role of MDM & Data Governance in the Digital Era. *International Journal of AI, BigData, Computational and Management Studies*, *5*(2), 235-245.

[20] Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, *3*, 160018.

[21] Aluri, Y. S. (2025). Frontend Performance Optimization of Large-Scale E-commerce Landing Pages: A Comprehensive Analysis. *International Journal of Computational and Experimental Science and Engineering*, *11*(4).

[22] Stall, S., Yarmey, L., Cutcher-Gershenfeld, J., Hanson, B., Lehnert, K., Nosek, B., ... & Wyborn, L. (2019). Make scientific data FAIR. Nature, *570*(7759), 27–29.

[23] Kumar, M. S., & Yuvaraj, N. (2024). Predictive Customer Experience Orchestration Using Governed Data Pipelines and Intelligent Service Signals. *International Journal of Emerging Trends in Computer Science and Information Technology*, *5*(1), 206-215.

[24] Boeckhout, M., Zielhuis, G. A., & Bredenoord, A. L. (2018). The FAIR guiding principles for data stewardship: Fair enough? European Journal of Human Genetics, *26*(7), 931–936.

[25] Kumar, M. S. (2023). A Scalable Architecture for Automated Data Classification and Sensitive Information Discovery Using Artificial Intelligence. *International Journal of Emerging Research in Engineering and Technology*, *4*(2), 158-169.

[26] European Commission. (2020). Turn FAIR data into reality. European Commission.

[27] GO FAIR. (2020). The FAIR Data principles. GO FAIR International Support and Coordination Office.

[28] Jaladi, D. S., & Mallempati, A. (2024). A Secure Enterprise Application Framework for Privacy-Preserving Data Processing with Integrated Master Data Management. *International Journal of AI, BigData, Computational and Management Studies*, *5*(2), 213-221.

[29] Lamprecht, A. L., Garcia, L., Kuzak, M., Martinez, C., Arcila, R., Martin Del Pico, E., ... & Gonzalez-Beltran, A. (2020). Towards FAIR principles for research software. Data Science, *3*(1), 37–59.

[30] Sansone, S. A., McQuilton, P., Rocca-Serra, P., Gonzalez-Beltran, A., Izzo, M., Lister, A. L., & Thurston, M. (2019). FAIRsharing as a community approach to standards, repositories and policies. Nature Biotechnology, *37*(4), 358–367.

[31] Jacobsen, A., de Miranda Azevedo, R., Juty, N., Batista, D., Coles, S., Cornet, R., ... & Schultes, E. (2020). FAIR principles: Interpretations and implementation considerations. Data Intelligence, *2*(1-2), 10–29.

[32] Batini, C., Cappiello, C., Francalanci, C., & Maurino, A. (2009). Methodologies for data quality assessment and improvement. ACM Computing Surveys, *41*(3), 1–52.

[33] Jaladi, D. S., & Mallempati, A. (2025). A Real-Time Enterprise Application Architecture for High-Volume Data Processing with Integrated Master Data Management. *International Journal of Emerging Research in Engineering and Technology*, *6*(1), 126-134.