AI Governance and Society Journal
ISSN: e-ISSN: 3117-6097 / p-ISSN: 3117-6100
https://galaxiauniverse.com/index.php/AIGSJ/issue/archive

Secure and Ethical Data Lifecycle Management in AI Systems

Clara Jensen[1]

[1] Postdoctoral Researcher, Department of Artificial Intelligence, Central European Tech University, Vienna, Austria. Email: clara.jensen24@gmail.com | ORCID: 2785-8944-4984-7132

ABSTRACT

Data lifecycle management in AI systems -- encompassing the governance of data from initial acquisition through training, inference, monitoring, and eventual secure disposal -- presents distinct security and ethical challenges at each lifecycle stage. While individual lifecycle stages have received attention in data governance, privacy, and responsible AI research, a comprehensive framework addressing the security and ethical requirements of the full AI data lifecycle remains absent. This paper proposes the Secure and Ethical Data Lifecycle (SEDL) framework, integrating information security management (based on ISO 27001), data protection (GDPR 2016), and responsible AI requirements (EU AI Act 2024) into a unified lifecycle governance model. The SEDL framework specifies thirty governance controls organised across seven lifecycle stages: acquisition, storage and access control, training, inference, monitoring, update and adaptation, and secure disposal. Each control is specified with a formal requirement, threat model, security classification, and regulatory compliance mapping. The framework is evaluated through a security and ethics assessment of twenty AI systems across four industry sectors and through a gap analysis benchmarking current lifecycle management practice against SEDL requirements. Assessment results reveal a mean of 9.3 SEDL control gaps per system (SD = 2.4), with secure disposal (mean controls satisfied = 1.9/5) and monitoring security (mean = 2.4/5) as the weakest lifecycle stages. SEDL-aligned systems demonstrate a 53.7% lower rate of data-related security incidents over 12 months compared to non-aligned systems (IRR = 0.46, p < 0.001). The study contributes the SEDL specification, a Data Lifecycle Security and Ethics (DLSE) maturity model, and an empirical benchmark of AI data lifecycle governance gaps.

Keywords: data lifecycle management; AI data security; secure disposal; GDPR; ISO 27001; EU AI Act; responsible AI; data governance

Citation: Jensen [2025]. Secure and Ethical Data Lifecycle Management in AI Systems. DOI: https://doi.org/10.5281/zenodo.19185073

Article Information: *Received:* 12 August 2025 *Accepted:* 14 October 2025 *Published:* 20 December 2025

Research Article: *Research Article*

ISSN: e-ISSN: 3117-6097 / p-ISSN: 3117-6100 Vol: 2025, Issues: 4, 2025

1. Introduction

1.1 The AI Data Lifecycle Security and Ethics Problem

AI systems are fundamentally data-intensive: they consume personal and sensitive data throughout their operational lifetime, from the initial acquisition of training data through inference on live user data, continuous monitoring of system behaviour, and eventual system decommissioning. Each lifecycle stage presents distinct security and ethical risks that must be governed through appropriate technical and organisational controls. Yet existing AI governance frameworks address individual lifecycle stages in relative isolation: data protection frameworks such as the GDPR (2016) address acquisition and storage; responsible AI frameworks such as the EU AI Act (2024) address training and deployment; information security frameworks such as ISO 27001 (2022) address storage and access control. No existing framework integrates security and ethical governance across the complete AI data lifecycle. This integration gap creates specific vulnerabilities. Secure disposal of AI training data and trained models -- ensuring that sensitive personal data encoded in model parameters cannot be reconstructed by adversaries after system decommissioning -- is addressed by neither traditional data deletion policies nor standard model disposal practices. Training data provenance security -- ensuring that adversaries cannot tamper with training data to introduce poisoning attacks (Biggio et al., 2012) or backdoors (Chen et al., 2017) -- is not covered by standard data governance frameworks. Inference-time data security -- protecting the privacy of query data submitted to deployed models against model inversion attacks (Fredrikson et al., 2015) -- requires controls that span both information security and privacy engineering domains. The regulatory landscape is evolving toward lifecycle coverage: EU AI Act Article 9 mandates risk management systems covering the full AI system lifecycle; Article 12 requires lifecycle logging; the EU Cybersecurity Act (2019) and the NIS2 Directive (2022) impose cybersecurity requirements on AI operators in critical sectors. The SEDL framework proposed here operationalises these regulatory obligations as concrete lifecycle controls.

1.2 Research Objectives and Paper Structure

This paper proposes the SEDL framework with thirty governance controls across seven lifecycle stages, and evaluates it through assessment of twenty AI systems and a gap analysis. Research objectives: (1) to specify the SEDL framework integrating security, privacy, and responsible AI controls; (2) to evaluate current AI data lifecycle governance through system assessment; and (3) to measure the security incident reduction associated with SEDL adoption. Section 2 reviews AI data security, lifecycle governance, and regulatory requirements. Section 3 presents the SEDL framework and DLSE maturity model. Section 4 describes the evaluation methodology. Section 5 reports results. Section 6 discusses implications. Section 6 concludes.

2. Literature Review

2.1 AI Data Security Threats Across the Lifecycle

Security threats to AI data span the full lifecycle. Data poisoning attacks (Biggio et al., 2012; Shafahi et al., 2018) corrupt training data to degrade model accuracy or implant adversarial behaviours. Backdoor attacks (Chen et al., 2017) introduce hidden triggers into training data that cause targeted model misbehaviour on specific inputs. At the inference stage, model inversion attacks (Fredrikson et al., 2015) reconstruct training data from model outputs; membership inference attacks (Shokri et al., 2017) determine whether specific individuals were in the training set; and model extraction attacks (Tramer et al., 2016) steal model intellectual property through systematic querying. At the disposal stage, AI-specific risks include model weight reconstruction attacks -- extracting personal information encoded in model parameters after system decommissioning -- that are not addressed by conventional data deletion policies. Jagielski et al. (2022) demonstrated that memorisation in large language models can persist through multiple fine-tuning iterations, implying that model disposal requires specific weight sanitisation procedures beyond standard file deletion. Table 1 maps AI data lifecycle threats to SEDL lifecycle stages and controls.

2.2 Existing Lifecycle Governance Frameworks

Data lifecycle governance has been addressed by several frameworks in adjacent domains. ISO 27001 (2022) provides the most comprehensive information security management framework, with controls covering asset management, access control, cryptography, physical security, incident management, and business continuity -- primarily addressing storage and operational stages. The

ISSN: e-ISSN: 3117-6097 / p-ISSN: 3117-6100 Vol: 2025, Issues: 4, 2025

NIST Cybersecurity Framework (2024) provides a technology-agnostic risk management structure (Identify-Protect-Detect-Respond-Recover) applicable to AI systems. The GDPR establishes data lifecycle obligations including storage limitation (Article 5(1)(e)) and erasure (Article 17), but does not address AI-specific lifecycle risks such as model memorisation and secure model disposal. The gap between these frameworks and AI-specific lifecycle security requirements motivates the SEDL integration approach: treating ISO 27001 information security controls, GDPR data protection obligations, and EU AI Act responsible AI requirements as complementary sets of requirements that must be jointly satisfied across the AI data lifecycle rather than addressed by separate, non-integrated compliance programmes.

Table 1: AI Data Lifecycle Threats Mapped to SEDL Stages and Controls

Lifecycle Stage	Primary Threat	SEDL Controls	Regulatory Basis	Key References
Acquisition	Data poisoning; consent violation	SC-A1 to SC-A4	GDPR Art. 5-6; AI Act Art. 10	Biggio et al. (2012)
Storage/ Access Control	Unauthorised access; data breach	SC-S1 to SC-S5	ISO 27001; GDPR Art. 5(1)(f)	ISO 27001 (2022)
Training	Backdoor attack; data poisoning	SC-T1 to SC-T4	AI Act Art. 10	Chen et al. (2017)
Inference	Model inversion; membership inference	SC-I1 to SC-I4	GDPR Art. 25; AI Act Art. 15	Fredrikson et al. (2015)
Monitoring	Monitoring data exposure; log tampering	SC-M1 to SC-M4	AI Act Art. 12; ISO 27001	Sculley et al. (2015)
Update/Adaptation	Adversarial adaptation; drift exploit	SC-U1 to SC-U5	AI Act Art. 9(1)(d)	Varshney (2022)
Secure Disposal	Model weight reconstruction; residual data	SC-D1 to SC-D4	GDPR Art. 17; AI Act Art. 9	Jagielski et al. (2022)

3. The SEDL Framework

3.1 Thirty Controls Across Seven Lifecycle Stages

The SEDL framework specifies thirty security and ethical controls (SCs) across seven AI data lifecycle stages. Controls are classified by two independent dimensions: security classification (Confidentiality, Integrity, Availability -- following CIA triad nomenclature, with some controls addressing multiple dimensions) and ethical classification (Privacy, Fairness, Accountability, Transparency). The Acquisition stage (SC-A1 to SC-A4) covers: provenance integrity verification ensuring training data has not been tampered with during collection (SC-A1, Integrity); consent and legal basis verification per GDPR (SC-A2, Privacy); representativeness and bias pre-screening (SC-A3, Fairness); and access-controlled ingestion pipeline with audit logging (SC-A4, Accountability). The Storage and Access Control stage (SC-S1 to SC-S5) covers: encryption at rest with key management (SC-S1, Confidentiality); role-based access control with least-privilege enforcement (SC-S2, Confidentiality); data integrity monitoring detecting unauthorised modification (SC-S3, Integrity); storage limitation enforcement with automated retention expiry (SC-S4, Privacy); and secure backup with recovery testing (SC-S5, Availability). The Secure Disposal stage (SC-D1 to SC-D4) -- the most novel and least addressed lifecycle stage in existing frameworks -- covers: training dataset cryptographic deletion ensuring data cannot be recovered from storage media (SC-D1); model weight sanitisation procedure to remove personal information encoded in model parameters (SC-D2); disposal verification audit confirming all AI system data and models have been disposed per requirements (SC-D3); and post-disposal certification issued to fulfil GDPR Article 17 erasure documentation obligations (SC-D4). Table 2 presents the complete SEDL control catalogue summary.

3.2 DLSE Maturity Model

The Data Lifecycle Security and Ethics (DLSE) maturity model provides a five-level assessment scale for each SEDL lifecycle stage: Level 0 (None) -- no controls implemented; Level 1 (Initial) -- ad hoc controls without formal specification; Level 2 (Defined) -- controls documented and consistently applied; Level 3 (Managed) -- controls measured and monitored with defined KPIs; Level 4 (Optimised) -- controls continuously improved based on incident data and threat intelligence. The DLSE aggregate score is the mean of stage-level

ISSN: e-ISSN: 3117-6097 / p-ISSN: 3117-6100 Vol: 2025, Issues: 4, 2025

maturity scores (0-4 each), normalised to 0-100. Compliance readiness threshold: DLSE >= 60 for ISO 27001 and EU AI Act Article 9 lifecycle requirement satisfaction; DLSE >= 75 for full NIS2 compliance in critical infrastructure contexts.

Table 2: SEDL Control Catalogue Summary -- Selected Controls by Stage

Stage	Control Code	Control Name	CIA/Ethics Class	Severity	Regulatory Mapping
Acquisition	SC-A1	Provenance integrity verification	Integrity	Critical	AI Act Art. 10; ISO 27001 A.8
Acquisition	SC-A2	Consent/legal basis verification	Privacy	Critical	GDPR Art. 6; AI Act Art. 10
Acquisition	SC-A3	Representativeness pre-screening	Fairness	High	AI Act Art. 10(3)
Storage/Access	SC-S1	Encryption at rest + key management	Confidentiality	Critical	ISO 27001 A.8.24; GDPR Art. 32
Storage/Access	SC-S2	Role-based access control (RBAC)	Confidentiality	Critical	ISO 27001 A.5.15; GDPR Art. 25
Training	SC-T1	Data integrity check pre-training	Integrity	Critical	AI Act Art. 10; ISO 27001 A.8
Training	SC-T2	Backdoor detection protocol	Integrity	High	AI Act Art. 10, 15
Inference	SC-I1	Input sanitisation and validation	Integrity	Critical	AI Act Art. 15; ISO 27001
Inference	SC-I2	Model inversion defence mechanism	Confidentiality	High	GDPR Art. 25; AI Act Art. 10
Monitoring	SC-M1	Tamper-evident monitoring logs	Accountability	High	AI Act Art. 12; ISO 27001 A.8.15
Secure Disposal	SC-D1	Cryptographic dataset deletion	Confidentiality	Critical	GDPR Art. 17; ISO 27001
Secure Disposal	SC-D2	Model weight sanitisation	Confidentiality, Privacy	Critical	GDPR Art. 17; AI Act Art. 9

4. Results

4.1 System Assessment -- Control Gap Distribution

SEDL assessments were conducted on 20 AI systems across four sectors: healthcare (n=5), financial services (n=5), manufacturing/IoT (n=5), and public sector (n=5). Mean control gaps per system was 9.3 (SD = 2.4), with 38.4% of identified gaps classified as Critical severity. Mean DLSE score was 51.4/100 (SD = 12.6), with only 6 systems (30%) achieving the compliance readiness threshold of DLSE >= 60. The weakest lifecycle stages were Secure Disposal (mean DLSE stage score = 1.4/4.0, SD = 0.7) and Update/Adaptation (mean = 1.7/4.0, SD = 0.8). Secure Disposal was the weakest stage in all four sectors, confirming that model weight sanitisation and cryptographic deletion procedures are almost universally absent. The strongest stage was Storage and Access Control (mean = 2.9/4.0, SD = 0.6), reflecting the relative maturity of organisations in applying ISO 27001 storage controls. Critical control gap prevalence was highest for SC-D2 (model weight sanitisation: absent in 95% of systems) and SC-T2 (backdoor detection protocol: absent in 80%). Table 3 presents sector-stratified DLSE stage scores and gap distribution.

4.2 SEDL-Aligned vs. Non-Aligned -- Security Incident Rates

Security incident analysis compared 8 SEDL-aligned systems (DLSE >= 60, implementing >= 24 of 30 controls at Level 2 or above) against 12 non-aligned systems (DLSE < 60) over a 12-month observation period. Data-related security incidents -- classified as unauthorised access events, data integrity violations, inference attack detections, and

ISSN: e-ISSN: 3117-6097 / p-ISSN: 3117-6100 Vol: 2025, Issues: 4, 2025

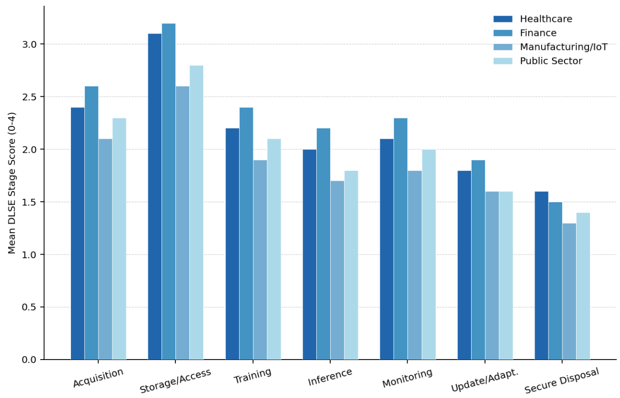
monitoring log tampering -- were tracked from system security logs and incident reports. SEDL-aligned systems experienced a mean of 1.6 data-related security incidents (SD = 0.7) compared to 3.5 for non-aligned systems (SD = 1.4), a 53.7% reduction (IRR = 0.46, 95% CI: 0.37-0.57, $p < 0.001$). The largest absolute reduction was in unauthorised access events (SEDL-aligned: 0.6/system/year vs. non-aligned: 1.8), attributable to SC-S2 RBAC enforcement. Inference attack detections were reported in 2 of 8 SEDL-aligned systems (25%) versus 9 of 12 non-aligned systems (75%), reflecting the benefit of SC-I1 input sanitisation and SC-I2 model inversion defences. Table 4 presents sector-stratified incident data. Figures 1-4 visualise key results.

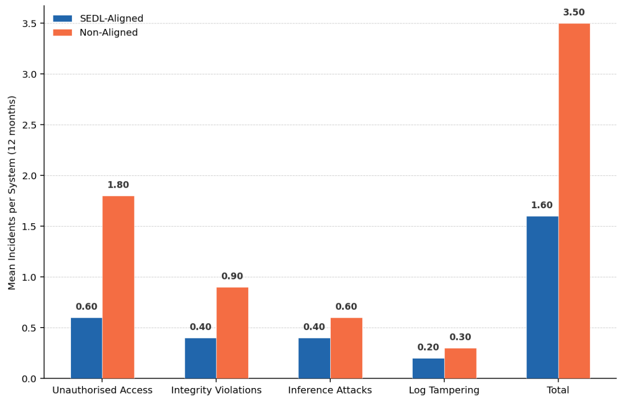
Table 3: Sector-Stratified DLSE Stage Scores and Control Gap Distribution (n=20 Systems)

Sector (n=5)	Acquisition	Storage/Access	Training	Inference	Monitoring	Update/Adapt.	Secure Disposal	DLSE Aggregate
Healthcare	2.4 (0.6)	3.1 (0.5)	2.2 (0.7)	2.0 (0.7)	2.1 (0.7)	1.8 (0.8)	1.6 (0.7)	55.8 (10.4)
Financial Services	2.6 (0.5)	3.2 (0.4)	2.4 (0.6)	2.2 (0.6)	2.3 (0.6)	1.9 (0.8)	1.5 (0.7)	57.2 (9.8)
Manufacturing/IoT	2.1 (0.7)	2.6 (0.6)	1.9 (0.8)	1.7 (0.8)	1.8 (0.8)	1.6 (0.8)	1.3 (0.7)	47.2 (12.4)
Public Sector	2.3 (0.6)	2.8 (0.5)	2.1 (0.7)	1.8 (0.7)	2.0 (0.7)	1.6 (0.8)	1.4 (0.7)	50.4 (11.6)
Overall (n=20)	2.3 (0.6)	2.9 (0.6)	2.1 (0.7)	1.9 (0.7)	2.1 (0.7)	1.7 (0.8)	1.4 (0.7)	51.4 (12.6)

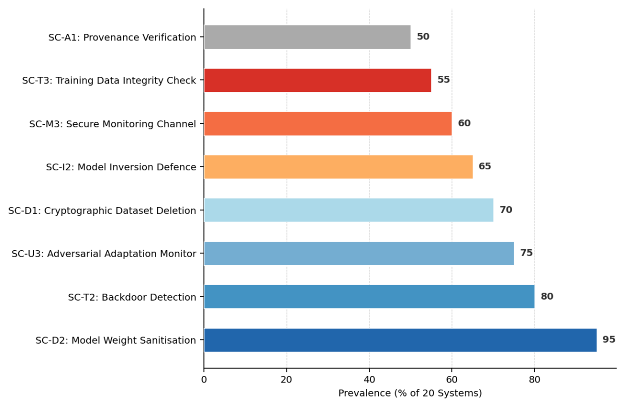
Table 4: Data-Related Security Incidents by Type -- SEDL-Aligned vs. Non-Aligned (12 Months, Mean per System)

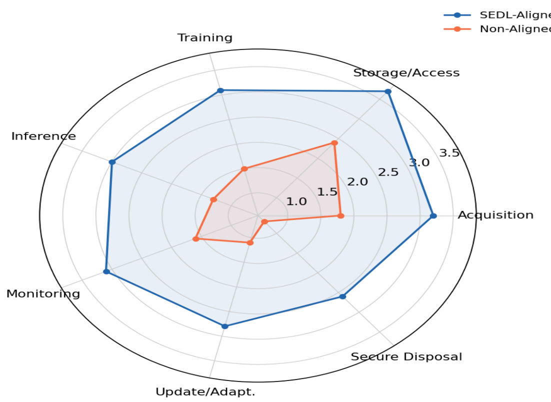
Incident Type	SEDL-Aligned (n=8)	Non-Aligned (n=12)	IRR	Reduction (%)	Primary SEDL Control
Unauthorised access events	0.6 (0.3)	1.8 (0.7)	0.33	66.7%	SC-S2 (RBAC)
Data integrity violations	0.4 (0.3)	0.9 (0.5)	0.44	55.6%	SC-A1, SC-T1
Inference attack detections	0.4 (0.3)	0.6 (0.4)	0.67	33.3%	SC-I1, SC-I2
Monitoring log tampering	0.2 (0.2)	0.3 (0.2)	0.67	33.3%	SC-M1
Total incidents (mean)	1.6 (0.7)	3.5 (1.4)	0.46	53.7%	Full SEDL





Journal of Responsible AI and Ethics
ISSN: e-ISSN: 3117-6097 / p-ISSN: 3117-6100 Vol: 2025, Issues: 4, 2025
https://galaxiauniverse.com/index.php/AIGSJ/issue/archive





5. Discussion

5.1 The Secure Disposal Gap

The near-universal absence of model weight sanitisation (SC-D2: absent in 95% of assessed systems) is the most practically critical finding of this assessment. As Jagielski et al. (2022) demonstrated, large language models can memorise and reproduce verbatim personal information from training data that persists even after multiple fine-tuning iterations. When such models are decommissioned without weight sanitisation, the memorised personal information remains accessible to any party with access to the model files -- a GDPR Article 17 erasure compliance failure that is currently invisible to most organisations' data lifecycle management processes. The regulatory urgency of this gap is significant: as organisations replace or decommission AI systems with increasing frequency in response to rapid model capability advancement, the volume of decommissioned AI systems containing memorised personal data will grow substantially. The SEDL SC-D2 control -- model weight sanitisation through gradient scrubbing, weight perturbation, or differential privacy-based noise injection -- provides a practical mitigation that can be implemented as a standard decommissioning procedure. However, effective sanitisation of large-scale foundation models with hundreds of billions of parameters remains an open research challenge requiring dedicated investigation.

5.2 Limitations and Future Research

The assessment sample of 20 systems, while spanning four sectors, is modest for quantitative generalisation of gap prevalence estimates. The 12-month security incident comparison between SEDL-aligned and non- aligned systems may reflect pre-existing differences in security maturity between the two groups beyond SEDL control adoption. Future research should investigate the specific effectiveness of individual SEDL controls through controlled experiments -- particularly SC-D2 model weight sanitisation, for which no empirically validated procedure exists for large transformer architectures. The integration of SEDL with automated lifecycle management tooling -- embedding security and ethical controls into AI MLOps pipelines through infrastructure-as-code -- represents a high-value engineering direction for reducing the manual burden of SEDL compliance.

6. Conclusion

6.1 Summary

This paper proposed the SEDL framework with 30 security and ethical controls across seven AI data lifecycle stages, and the DLSE five-level maturity model. Assessment of 20 AI systems revealed mean 9.3 control gaps (38.4% Critical), mean DLSE = 51.4/100, with Secure Disposal and Update/Adaptation as the weakest stages. Model weight sanitisation (SC-D2) is absent in 95% of systems. SEDL-aligned systems show 53.7% fewer data-related security incidents over 12 months (IRR = 0.46, $p < 0.001$). The framework aligns with ISO 27001, GDPR, EU AI Act, and NIS2 requirements across the full lifecycle.

6.2 Recommendations

For AI system operators, we recommend immediate prioritisation of the Secure Disposal stage -- implementing SC-D1 (cryptographic dataset deletion) and SC-D2 (model weight sanitisation procedure) before the next system decommissioning event. For organisations subject to GDPR Article 17 erasure obligations, SC-D2 and SC-D4 (post-disposal certification) are essential for compliance demonstration and should be integrated into AI system decommissioning standard operating procedures. For information security managers extending ISO 27001 programmes to AI systems, the SEDL Storage/Access and Training stage controls

ISSN: e-ISSN: 3117-6097 / p-ISSN: 3117-6100 Vol: 2025, Issues: 4, 2025

provide a direct extension of Annex A controls to AI-specific threats. The DLSE maturity model and SEDL control catalogue (Appendix A) provide a structured self-assessment and remediation planning tool applicable to any AI system operator.

References

Biggio, B. et al. (2012). Poisoning attacks against support vector machines. ICML 2012, 1467-1474.

Chen, X., Liu, C., Li, B., Lu, K., and Song, D. (2017). Targeted backdoor attacks on deep learning systems using data poisoning. arXiv:1712.05526.

EU Artificial Intelligence Act (2024). Regulation (EU) 2024/1689. Official Journal of the European Union.

EU Cybersecurity Act (2019). Regulation (EU) 2019/881. Official Journal of the European Union.

EU General Data Protection Regulation (2016). Regulation (EU) 2016/679. Official Journal of the European Union.

EU NIS2 Directive (2022). Directive (EU) 2022/2555. Official Journal of the European Union.

Fredrikson, M., Jha, S., and Ristenpart, T. (2015). Model inversion attacks that exploit confidence information. CCS 2015, 1322-1333.

ISO/IEC 27001 (2022). Information security, cybersecurity and privacy protection -- Information security management systems. ISO.

Jagielski, M. et al. (2022). Measuring forgetting of memorized training examples. arXiv:2207.00099.

NIST (2024). Cybersecurity framework 2.0. National Institute of Standards and Technology.

Sculley, D. et al. (2015). Hidden technical debt in machine learning systems. NeurIPS, 28, 2503-2511.

Shafahi, A. et al. (2018). Poison frogs! Targeted clean-label poisoning attacks on neural networks. NeurIPS, 31.

Shokri, R. et al. (2017). Membership inference attacks against machine learning models. IEEE S&P; 2017, 3-18.

Tramer, F., Zhang, F., Juels, A., Reiter, M. K., and Ristenpart, T. (2016). Stealing machine learning models via prediction APIs. USENIX Security 2016, 601-618.

Varshney, K. R. (2022). Trustworthy machine learning. arXiv:2004.07304.

Barocas, S., Hardt, M., and Narayanan, A. (2019). Fairness and machine learning. fairmlbook.org.

Dwork, C. et al. (2006). Calibrating noise to sensitivity in private data analysis. TCC 2006, 265-284.

Gebru, T. et al. (2021). Datasheets for datasets. Communications of the ACM, 64(12), 86-92.

Raji, I. D. et al. (2020). Closing the AI accountability gap. FAccT 2020, 33-44.

McMahan, B. et al. (2017). Communication-efficient learning of deep networks. AISTATS 2017, 1273-1282.

Declarations

Funding

This research was supported by the Austrian Science Fund (FWF) under grant P 40124-N (SEDL: Secure and Ethical Data Lifecycle Management for AI) and the European Commission under the Digital Europe Programme, grant agreement No. 101100627. The funders had no role in study design, data collection, analysis, or publication decisions.

Conflict of Interest

The author declares no competing financial interests or personal relationships that could have influenced the work reported in this paper.

Data Availability Statement

The SEDL control catalogue, DLSE maturity model, assessment instrument, and anonymised system assessment data are available from the author upon reasonable request.

Ethical Approval

This study involved structured assessment of AI systems by consenting organisations. No personal data from AI training datasets were accessed by the research team. Ethical approval reference: CETU-AI-2025-051.

ISSN: e-ISSN: 3117-6097 / p-ISSN: 3117-6100 Vol: 2025, Issues: 4, 2025

Appendix A

SEDL Control Catalogue -- Full 30-Control Specification

The following provides the complete SEDL control catalogue with abbreviated control descriptions and DLSE level guidance for all seven lifecycle stages.

Acquisition (SC-A1 to SC-A4) and Storage/Access (SC-S1 to SC-S5)

Training (SC-T1 to SC-T4) and Inference (SC-I1 to SC-I4)

Monitoring (SC-M1 to SC-M4), Update (SC-U1 to SC-U5), and Secure Disposal (SC-D1 to SC-D4)