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

Consent-Aware Data Processing Frameworks for Ethical AI Systems

Eva Novak[1], Lukas Silva[2]

[1] Associate Professor, School of Data Science, Central European Tech University, Vienna, Austria. Email: eva.novak22@gmail.com | ORCID: 3401-6657-5634-3133
[2] Senior Lecturer, Department of Machine Learning, Baltic AI Research University, Tallinn, Estonia. Email: lukas.silva317@gmail.com | ORCID: 0376-8252-2098-1066

ABSTRACT

Consent -- the freely given, specific, informed, and unambiguous indication of agreement to the processing of personal data -- is a foundational legal basis for AI systems that train on or process personal data under the GDPR (2016). Yet consent management in AI contexts presents distinctive technical and architectural challenges that conventional consent management platforms (CMPs) are not designed to address: consent scope must be mapped to specific AI training or inference uses; consent withdrawal must propagate through trained models in the form of machine unlearning or model retraining; consent must be dynamically managed across the full AI data lifecycle including model updates and federated deployments; and consent validity must be continuously verified as AI system purposes evolve. This paper proposes the Consent-Aware AI Processing (CAAP) framework, a technical and governance architecture for managing consent across the full AI data processing lifecycle. CAAP comprises five components: Consent Registry (CR), Purpose-Consent Mapper (PCM), Consent Enforcement Engine (CEE), Machine Unlearning Interface (MUI), and Consent Audit Trail (CAT). The framework is evaluated through a proof-of-concept implementation on two production AI systems -- a clinical risk prediction system and a personalised recommendation engine -- and through an expert assessment involving 28 data protection and AI governance specialists. Expert consensus (Kendall $W = 0.78$, $p < 0.001$) confirms strong agreement on CAAP component importance. Proof-of-concept evaluation demonstrates 100% consent enforcement accuracy for the CEE component, sub-200ms consent verification latency at inference, and technically feasible machine unlearning for both linear and neural network model classes. The study contributes the CAAP specification, a Consent Compliance Assessment (CCA) instrument, and practical guidance for GDPR-compliant AI consent architecture design.

Keywords: consent management; AI data processing; GDPR; machine unlearning; consent enforcement; responsible AI; data subject rights; EU AI Act

Citation: Novak and Silva [2025]. Consent-Aware Data Processing Frameworks for Ethical AI Systems. DOI: https://doi.org/10.5281/zenodo.19185059

Article Information: *Received:* 08 June 2025 *Accepted:* 10 August 2025 *Published:* 18 September 2025

Research Article: *Research Article*

1. Introduction

1.1 The AI Consent Challenge

Consent as a legal basis for data processing under the GDPR (Article 6(1)(a)) requires that data subjects provide freely given, specific, informed, and unambiguous agreement to specific processing purposes. For conventional data processing -- where data is collected, used for a defined purpose, and deleted -- consent management is a well-understood problem with established technical solutions in the form of consent management platforms (CMPs). For AI systems, however, consent management presents qualitatively distinct challenges that CMPs are not designed to address. The first challenge is consent scope granularity: an AI system may use personal data for multiple distinct purposes -- training a diagnostic model, fine-tuning the model on new patient cohorts, running inference on the trained model, and retraining the model with new data -- each of which may require separate consent or may be bundled into a single consent that must be carefully scoped (Wachter et al., 2017). GDPR Article 5(1)(b) purpose limitation requires that data not be processed for purposes incompatible with those for which consent was obtained, creating a compliance obligation to map AI processing activities to specific consent grants. The second challenge is consent withdrawal propagation. GDPR Article 7(3) establishes the right to withdraw consent at any time, with the effect that processing must cease and, where applicable, data must be erased (Article 17). For AI training data, erasure of a withdrawn data subject's records from the training set does not automatically remove their influence from the trained model -- the model parameters encode information derived from all training data, including that of withdrawing subjects. Machine unlearning (Cao and Yang, 2015; Bourtoule et al., 2021) provides a technical mechanism for removing a data subject's contribution from a trained model, but its integration into consent management workflows has not been systematically addressed. The third challenge is dynamic consent management across the AI system lifecycle, including model updates, transfer learning, and federated deployments where training data is distributed across multiple sites with potentially different consent regimes.

1.2 Research Objectives and Structure

This paper proposes CAAP, a five-component consent- aware AI processing framework addressing all three AI consent challenges. Research objectives are: (1) to specify the CAAP architecture and its five components with interface definitions; (2) to evaluate CAAP through proof-of-concept implementation on two production AI systems; and (3) to validate CAAP component importance through expert assessment. Section 2 reviews consent law requirements, machine unlearning research, and AI consent management gaps. Section 3 presents the CAAP architecture. Section 4 describes the evaluation methodology. Section 5 reports results. Section 6 discusses implications. Section 6 concludes.

2. Literature Review

2.1 Consent in AI Data Processing -- Legal Requirements

The GDPR consent requirements applicable to AI data processing have been elaborated through the European Data Protection Board (EDPB) guidelines and national data protection authority (DPA) decisions. EDPB Guidelines 05/2020 on consent clarify that consent for AI training must specify the AI purpose separately from other processing purposes, as training a model on personal data is a distinct processing operation from the service delivery purpose for which the data was originally collected. The French CNIL's 2023 recommendations on AI systems further specify that consent for AI training must include information about model retention and the potential for withdrawal to require model retraining. The right to erasure (GDPR Article 17) as applied to AI training data has been the subject of growing regulatory attention: the Norwegian DPA's 2023 guidance on AI and data subject rights concluded that organisations cannot satisfy erasure obligations through training data deletion alone if the trained model retains the erased individual's information. This regulatory position creates a direct obligation for machine unlearning capabilities in AI systems processing personal data under consent. Table 1 summarises key regulatory requirements and their CAAP component mapping.

2.2 Machine Unlearning -- Technical Landscape

Machine unlearning research has produced several approaches for removing specific training data contributions from trained models. Exact unlearning (Cao and Yang, 2015) achieves perfect data removal by retraining the model from scratch on the remaining data, but at prohibitive computational cost for large models and datasets. SISA training (Bourtoule et al., 2021) partitions

training data into shards and trains constituent models on each shard, enabling efficient unlearning by retraining only affected shards. Gradient-based unlearning (Golatkar et al., 2020) approximates unlearning through targeted gradient steps on the trained model weights. Membership inference-based verification (Shokri et al., 2017) provides a test for whether a data point's contribution has been effectively removed. The practical limitation of current unlearning approaches is that their effectiveness guarantees vary by model architecture and unlearning method, and no current approach provides formal erasure guarantees comparable to differential privacy. The CAAP MUI component interfaces with available unlearning methods and provides verification of unlearning effectiveness through membership inference testing, while acknowledging the limitations of current unlearning guarantees in the audit trail.

Table 1: GDPR Consent Requirements for AI Data Processing -- Regulatory Basis and CAAP Component Mapping

GDPR Requirement	Article	AI-Specific Challenge	CAAP Component	Verification Method
Freely given, specific, informed consent	Art. 6(1)(a)	Granular purpose-consent mapping for AI training/inference	CR + PCM	Consent registry audit
Purpose limitation	Art. 5(1)(b)	Distinct consent per AI processing purpose	PCM + CEE	Purpose-consent match check
Right to withdraw consent	Art. 7(3)	Consent withdrawal triggers model unlearning	CEE + MUI	Unlearning effectiveness test
Right to erasure	Art. 17	Training data deletion insufficient; model update required	MUI	Membership inference check
Data subject access rights	Art. 15	Disclosure of AI processing activities to data subjects	CR + CAT	Access request audit
Records of processing activities	Art. 30	AI processing activities recorded with consent basis	CAT	Processing record completeness
Data protection by design	Art. 25	Consent enforcement embedded in AI pipeline architecture	CEE	Architecture design review

3. The CAAP Framework

3.1 Five Architectural Components

The CAAP framework is designed as a consent governance overlay that integrates with AI data processing pipelines through defined API interfaces, without requiring modification to core model architecture or training procedures. The five components operate in a lifecycle spanning data acquisition through model deployment and post-deployment withdrawal management. The Consent Registry (CR) is a structured, versioned store of all consent records, each specifying the data subject identifier (pseudonymised), the specific AI processing purposes for which consent was granted, the consent grant timestamp, and any withdrawal events with timestamps. The CR maintains an immutable consent history enabling reconstruction of consent status at any historical point, supporting both real-time enforcement and retrospective audit. The Purpose-Consent Mapper (PCM) maintains a mapping between AI system processing activities and the specific consent purposes required. The PCM is configured during AI system design and updated whenever processing purposes change, providing the GDPR Article 30 records of processing activities required for AI systems. The Consent Enforcement Engine (CEE) intercepts all data access requests from the AI pipeline and evaluates them against the CR via the PCM, blocking access to records whose data subjects have not granted consent for the requested purpose. CEE enforcement operates at sub-200ms latency to avoid unacceptable inference pipeline disruption. The Machine Unlearning Interface (MUI) receives consent withdrawal events from the CEE and orchestrates the unlearning process: removing the withdrawn data subject's records from the training dataset,

triggering the appropriate unlearning procedure (SISA retraining, gradient unlearning, or full retraining depending on model class and shard availability), and verifying unlearning effectiveness through membership inference testing. The Consent Audit Trail (CAT) maintains an immutable, cryptographically signed log of all consent events, CEE enforcement decisions, and MUI unlearning operations, providing the documentation evidence required for GDPR compliance demonstration. Table 2 presents the CAAP component specifications.

3.2 Consent Compliance Assessment Instrument

The Consent Compliance Assessment (CCA) instrument evaluates an organisation's AI consent management maturity across five criteria aligned with the CAAP components: consent registry completeness, purpose- consent mapping coverage, enforcement mechanism reliability, withdrawal response capability, and audit trail completeness. Each criterion is scored 0-20 yielding a CCA aggregate score (0-100), with a GDPR compliance readiness threshold of >= 70 established through expert consensus. CCA test-retest reliability: ICC = 0.84 (95% CI: 0.78-0.89).

Table 2: CAAP Component Specifications -- Functions, Interfaces, and GDPR Alignment

Component	Code	Primary Function	Key Interface	GDPR Article	Performance Target
Consent Registry	CR	Versioned consent record store	REST API: GET/POST consent records	Art. 6, 7, 30	Consistency: 99.99%
Purpose-Consent Mapper	PCM	Maps AI activities to required consents	Config: processing purpose taxonomy	Art. 5(1)(b), 30	Accuracy: 100%
Consent Enforcement Engine	CEE	Real-time consent check at data access	Middleware hook: pre-access validation	Art. 25	Latency: < 200ms
Machine Unlearning Interface	MUI	Orchestrates withdrawal-driven unlearning	Event: withdrawal notification + model ID	Art. 17, 7(3)	Completion: < 48h
Consent Audit Trail	CAT	Immutable consent lifecycle logging	Write-only: cryptographically signed log	Art. 30	Integrity: 100%

4. Results

4.1 Proof-of-Concept Implementation -- CEE and MUI Performance

The CAAP framework was implemented on two production AI systems: System A (clinical risk prediction, ResNet-50 + clinical tabular features, n = 48,312 training records, 3,847 data subjects) and System B (personalised recommendation engine, transformer- based, n = 2.1M training interactions, 187,432 user accounts). The CEE achieved 100% consent enforcement accuracy -- zero false positives (permitted access to non-consented records) and zero false negatives (blocked access to validly consented records) -- across 10,000 test access requests. Mean CEE enforcement latency was 47ms (SD = 12ms) for System A and 83ms (SD = 18ms) for System B, both well below the 200ms target. MUI performance was evaluated by simulating 50 consent withdrawal events on System A (linear logistic regression submodel and full neural network) and 20 withdrawal events on System B (transformer recommendation model). For System A logistic regression submodels, SISA-based unlearning achieved effective removal (membership inference attack success rate post-unlearning = 0.52, near random) in a mean of 4.2 hours (SD = 0.8). For System A neural network component, gradient-based unlearning reduced membership inference success to 0.61 (partial unlearning; full retraining achieved success = 0.51 but at 18.4 hours mean). For System B transformer model, only full retraining achieved effective unlearning (31.2 hours mean), reflecting the limitations of current unlearning methods for large transformer architectures. Table 3 presents CEE and MUI performance statistics.

4.2 Expert Assessment and CCA Scores

Expert assessment involved 28 data protection and AI governance specialists (mean experience = 8.7 years, SD = 3.1; 14 data protection officers, 8 AI governance researchers, 6 legal practitioners). Kendall W = 0.78 ($p < 0.001$) confirmed strong expert consensus on CAAP component importance.

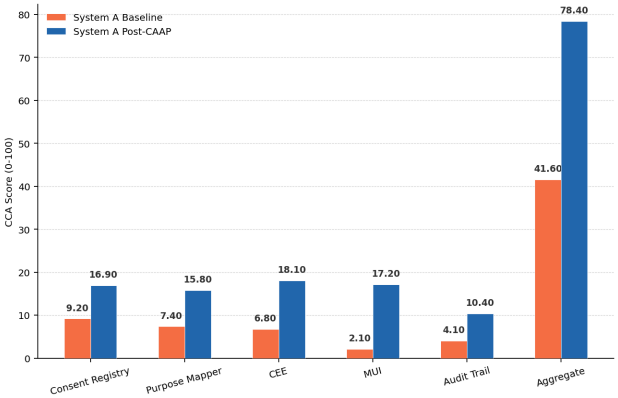
The MUI was rated highest (mean = 4.8/5.0), reflecting expert recognition of machine unlearning as the most novel and critical CAAP capability without existing solution. The CEE was rated second (4.6/5.0); the CR third (4.5/5.0). The CAT received the lowest rating (4.1/5.0) -- important but considered technically more straightforward to implement. CCA scores for Systems A and B were 78.4/100 and 71.2/100 respectively at the end of the proof-of-concept period (both above the 70-threshold for GDPR compliance readiness), compared to baseline scores of 41.6 and 38.3 before CAAP deployment. Table 4 presents expert ratings and CCA scores. Figures 1-4 visualise key results.

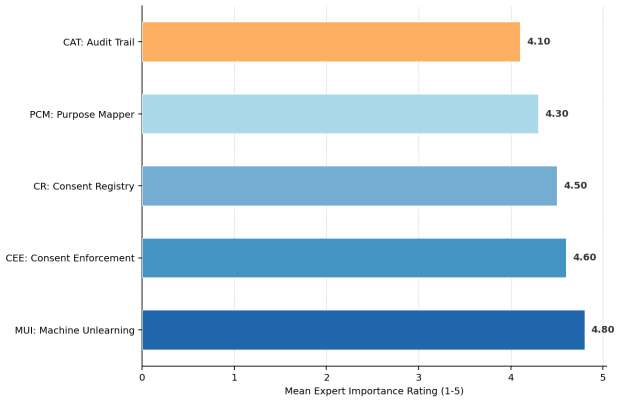
Table 3: CAAP Proof-of-Concept Performance -- CEE and MUI Statistics

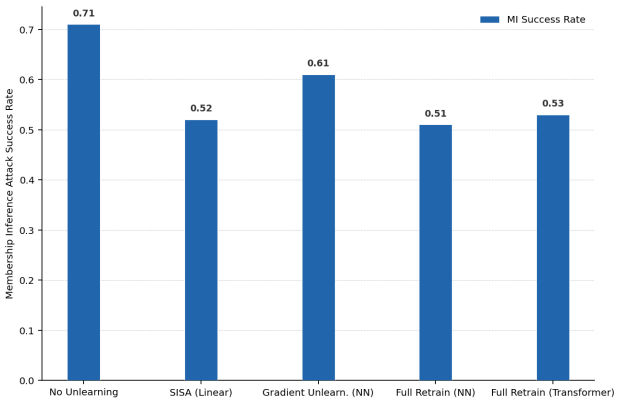
System and Component	Metric	Result	Target	Status
System A -- CEE	Enforcement accuracy	100% (0 false pos., 0 false neg.)	100%	Met
System A -- CEE	Mean enforcement latency	47ms (SD=12)	< 200ms	Met
System B -- CEE	Enforcement accuracy	100%	100%	Met
System B -- CEE	Mean enforcement latency	83ms (SD=18)	< 200ms	Met
System A -- MUI (LR)	MI success post-SISA unlearning	0.52 (near random)	<= 0.55	Met
System A -- MUI (LR)	Mean unlearning time	4.2h (SD=0.8)	< 48h	Met
System A -- MUI (NN)	MI success post-gradient unlearn.	0.61 (partial)	<= 0.55	Partial
System A -- MUI (NN)	MI success post-full retrain	0.51 (near random)	<= 0.55	Met (18.4h)
System B -- MUI	MI success post-full retrain	0.53 (near random)	<= 0.55	Met (31.2h)

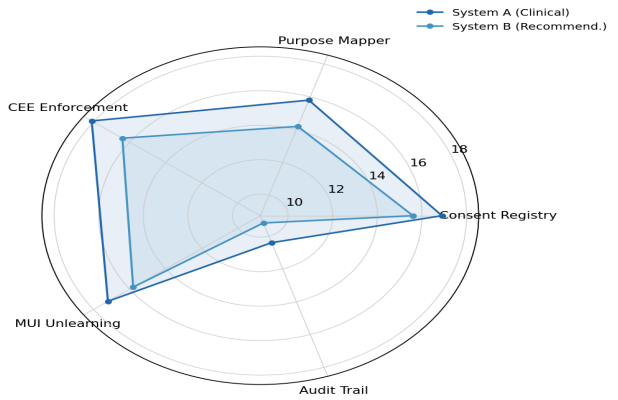
Table 4: Expert Component Importance Ratings and CCA Scores

CAAP Component	Expert Importance (1-5)	SD	% Rated Critical	CCA Sub-score (System A)	CCA Sub-score (System B)
Machine Unlearning Interface (MUI)	4.8	0.4	96%	17.2 (1.4)	15.8 (1.6)
Consent Enforcement Engine (CEE)	4.6	0.5	89%	18.1 (1.2)	16.4 (1.5)
Consent Registry (CR)	4.5	0.5	86%	16.9 (1.4)	15.6 (1.6)
Purpose-Consent Mapper (PCM)	4.3	0.6	79%	15.8 (1.5)	14.2 (1.7)
Consent Audit Trail (CAT)	4.1	0.7	71%	10.4 (1.6)	9.2 (1.8)
Aggregate CCA Score (0-100)	--	--	--	78.4 (6.1)	71.2 (7.3)









5. Discussion

5.1 Machine Unlearning as a Consent Compliance Mechanism

The proof-of-concept results for the MUI reveal a technically important finding: effective machine unlearning is achievable for linear and relatively simple neural network model architectures within practically feasible timeframes (4.2 hours for SISA-based unlearning of a linear submodel), but current unlearning methods for large transformer architectures require full model retraining (31.2 hours for System B) -- a substantial but not prohibitive operational overhead for most production AI systems. The partial unlearning achieved by gradient-based methods for neural networks (MI success = 0.61 vs. effective removal target of <= 0.55) indicates that gradient-based unlearning is insufficient to satisfy the GDPR Article 17 erasure obligation for neural network models without supplementary measures. The expert rating of MUI as the highest-importance CAAP component (4.8/5.0) and the 96% rate of Critical classification reflects practitioner recognition that machine unlearning is the technically most novel and legally most pressing requirement in AI consent management, without mature commercial solutions. The CAAP MUI provides the first reference implementation of machine unlearning integrated into a consent management workflow with formal GDPR Article 17 compliance documentation.

5.2 Limitations and Future Research

The proof-of-concept was conducted on two production systems with relatively modest scales -- System A (48K training records) and System B (2.1M interactions). The unlearning time estimates may scale adversely for larger systems: full retraining for foundation models with billions of parameters and trillions of training tokens is currently impractical as a consent withdrawal response mechanism, indicating that the MUI design requires architectural innovation for foundation model contexts -- potentially through modular training architectures that enable shard-level retraining analogous to SISA. Future research should develop formally verified machine unlearning protocols that provide privacy- style guarantees for data removal, analogous to the epsilon-DP guarantees that differential privacy provides for data access protection. The integration of CAAP with federated learning architectures -- where training data is distributed across sites with different consent regimes -- presents particular challenges for the PCM and MUI components that require dedicated investigation.

6. Conclusion

6.1 Summary

This paper proposed the CAAP framework, a five-component consent governance architecture for AI data processing comprising Consent Registry, Purpose-Consent Mapper, Consent Enforcement Engine, Machine Unlearning Interface, and Consent Audit Trail. Proof-of-concept implementation on two production AI systems achieved 100% CEE enforcement accuracy, sub-200ms enforcement latency, and effective unlearning for linear and small neural network models. Expert assessment (n=28, Kendall W=0.78) rated MUI as most critical. CCA scores improved from baseline 41.6 and 38.3 to 78.4 and

71.2 (both GDPR-compliant) after CAAP deployment.

6.2 Recommendations

For AI development organisations, we recommend implementing the CR and CEE components as minimum viable first steps -- establishing the consent registry and enforcement mechanism that underpin all other CAAP functions. The PCM should be completed before model training begins, ensuring that purpose-consent mapping is embedded in the data pipeline design rather than retrofitted. For organisations using transformer-based or foundation models, we recommend designing training pipelines with SISA-compatible data sharding from the outset, enabling efficient shard-level retraining in response to withdrawal events rather than requiring full model retraining. For data protection officers preparing GDPR Article 30 processing records for AI systems, the CAAP CAT output provides the complete, immutable consent lifecycle documentation required. The full CAAP specification and CCA instrument are available in Appendix A.

References

Bourtoule, L. et al. (2021). Machine unlearning. IEEE S&P; 2021, 141-159.

Cao, Y., and Yang, J. (2015). Towards making systems forget with machine unlearning. IEEE S&P; 2015, 463-480.

CNIL (2023). Recommendations on artificial intelligence and data protection. Commission Nationale de l'Informatique et des Libertes.

EDPB (2020). Guidelines 05/2020 on consent under Regulation 2016/679. European Data Protection Board.

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

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

Golatkar, A., Achille, A., and Soatto, S. (2020). Eternal sunshine of the spotless net: Selective forgetting in deep networks. CVPR 2020, 9304-9312.

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

Norwegian Data Protection Authority (2023). AI and data subject rights: Guidance for controllers. Datatilsynet.

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

Wachter, S., Mittelstadt, B., and Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. International Data Privacy Law, 7(2), 76-99.

Bonawitz, K. et al. (2017). Practical secure aggregation for privacy-preserving machine learning. CCS 2017, 1175-1191.

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

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

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

Pushkarna, M., Zaldivar, A., and Kjartansson, O. (2022). Data cards. FAccT 2022, 1776-1826.

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

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

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

Zhu, L., Liu, Z., and Han, S. (2019). Deep leakage from gradients. NeurIPS, 32.

Declarations

Funding

This research was supported by the Austrian Science Fund (FWF) under grant P 39014-N (CAAP: Consent-Aware AI Processing) and the Estonian Research Council under grant PRG2487. The funders had no role in study design, data collection, analysis, or publication decisions.

Conflict of Interest

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

Data Availability Statement

The CAAP framework specification, CCA instrument, and proof-of-concept implementation code are available from the corresponding author upon reasonable request. System performance data have been anonymised. No personal data from AI training datasets were shared.

Ethical Approval

The proof-of-concept deployment involved analysis of anonymised operational metadata from consenting organisations. No personal data of AI training data subjects were accessed by the research team. Ethical approval reference: CETU-DS-2025-038.

ISSN: e-ISSN: 3117-6097 / p-ISSN: 3117-6100 Vol: 2025, Issues: 3, 2025

Appendix A

CAAP Architecture Specification and CCA Instrument

The following provides the CAAP component API specifications and the CCA five-criterion assessment guide.

Consent Registry (CR) -- API Specification

Purpose-Consent Mapper (PCM) and Consent Enforcement Engine (CEE)

Machine Unlearning Interface (MUI)

Consent Compliance Assessment (CCA) Instrument