User Data Deletion and Retention Policy

Synthetic Users

Version: 1.2

Effective Date: March 2024

Last Updated: March 25, 2026

Document Owner: CTO — Artur Ventura

Review Frequency: Annually

Classification: Internal – Confidential

CRA Reference: 4.2.1, 18.1.2

Change History

Version	Date	Author	Changes
1.1	March 2024	CTO / Security Lead	Initial release
1.2	March 25, 2026	Artur Ventura, CTO	Added version metadata and last review date; added Section 16 (AI/GenAI Data Handling) covering embeddings, RAG pipeline data, model interaction logs, and decommissioning alignment; added related documents. Per JPMC SCA CRA 4.2.1 and 18.1.2.

1. Purpose

This policy defines how Synthetic Users manages, retains, and permanently deletes user data collected and processed on behalf of the organization and its clients. It ensures compliance with applicable data protection laws (including GDPR and CCPA), supports the organization's privacy commitments, and meets JPMC contractual data management obligations.

2. Data Mapping and Inventory

- Synthetic Users maintains a complete and up-to-date inventory of all data processed on behalf of the organization.
- Each dataset is classified according to sensitivity (personal, financial, confidential, operational).
- Data flows, storage locations, and transfer points are documented and reviewed at least annually.

3. Data Retention Policy

Synthetic Users maintains a documented data retention schedule specifying:

- Categories of data collected
- Purpose of processing
- Legal, contractual, and operational retention periods

Personal data is not retained longer than necessary for its original purpose. Upon expiry of the retention period, data is securely deleted or anonymized. Any exceptions (e.g., legal hold, dispute, or audit) are documented and justified.

4. Data Deletion Requests

- Synthetic Users maintains a dedicated channel (email or web form) for receiving deletion requests from clients or data subjects.
- Clear instructions for submitting deletion requests are made available.
- All deletion requests are logged and acknowledged within **five (5) business days**.

5. Verification of Requests

- A robust identity verification process is used to confirm the legitimacy of deletion requests and prevent unauthorized actions.
- Verification records are retained for audit purposes.

6. Secure Data Deletion

Valid deletion requests trigger **irreversible deletion** using secure industry-standard methods including cryptographic erasure or data shredding. Data is deleted from:

- Active systems and databases
- Logs and data lakes
- Replicated or cached environments
- Backups, following backup lifecycle policies (see Section 8)

Synthetic Users ensures data is not recoverable post-deletion.

7. Deletion Timeline

All valid data deletion requests are completed within **30 calendar days**, unless legal or regulatory requirements specify otherwise. The requester receives notification of completion or justification for any delays.

8. Backup and Archived Data

Personal data in backup or archived systems is either:

- Deleted immediately upon restoration, or
- Automatically purged during the next scheduled backup rotation (maximum retention: **90 days**)

Backups containing deleted data are encrypted and access-restricted until their deletion.

9. Third-Party Processors

- All subcontractors and subprocessors follow equivalent data retention and deletion standards.
- Contracts explicitly define data deletion obligations and audit rights.
- Synthetic Users remains fully accountable for ensuring subprocessor compliance.

10. Confirmation and Reporting

Upon completion of a deletion request, Synthetic Users issues written confirmation specifying:

- The data deleted
- The systems affected
- The date and method of deletion

Summary reports of deletion activities are available upon request.

11. Auditing and Compliance

- Internal audits are conducted at least annually to ensure compliance with this policy.
- Audit reports are retained and made available upon request or during regulatory inspections.
- Non-compliance incidents are reported within **72 hours** of discovery.

12. Employee Training

All employees and contractors with data access receive annual training on data retention, deletion procedures, and privacy best practices. Training completion is tracked and documented per the User Awareness Training Program.

13. Physical Media and Hardware Disposal

Any physical storage media (hard drives, USBs, servers) containing user data is destroyed using secure disposal methods such as degaussing or physical shredding. Disposal actions are logged and certified.

14. Continuous Improvement

This policy is reviewed and updated at least annually in line with evolving regulations, audit results, and best practices. Updates are communicated promptly to affected stakeholders.

15. Recordkeeping

Comprehensive records of all data deletion and retention actions are maintained for **a minimum of three (3) years** for audit and compliance verification.

16. AI/GenAI Data Handling

Synthetic Users processes data through AI/GenAI systems including LLM orchestration, the Persona Engine, and a RAG (Retrieval-Augmented Generation) pipeline. This section defines retention and deletion obligations specific to AI/GenAI data artifacts.

16.1 AI/GenAI Data Categories Subject to This Policy

Data Category	Description	Retention Period
User prompt inputs	Text submitted by users to AI-powered features	Session duration; not persistently stored
Model outputs / responses	AI-generated responses returned to users	Session duration; not persistently stored beyond session logs
RAG embeddings	Vector representations of tenant data stored in the retrieval pipeline	Retained for duration of active client contract; deleted within 30 days of contract termination

Model interaction logs	System-level logs recording prompt/response metadata (not full content)	90 days rolling, then purged
Evaluation and fine-tuning data	Data sets used to evaluate or improve model behavior	Retained for duration of active use; deleted upon project decommission per the AI/GenAI Decommissioning Policy
LLM provider data	Data sent to third-party LLM providers (OpenAI, Anthropic, Google, etc.)	Governed by provider DPA; not retained by provider beyond API response per DPA terms

16.2 Tenant Data Isolation

RAG pipeline data and embeddings are stored with tenant-scoped isolation. Deletion of a tenant's data removes all associated embeddings, retrieval indexes, and cached context from the RAG pipeline within the 30-calendar-day deletion timeline defined in Section 7.

16.3 LLM Provider Data Governance

Synthetic Users sends data to third-party LLM providers solely for inference (generating responses). Providers are contractually bound via Data Processing Agreements (DPAs) to:

- Not retain submitted data beyond the API request/response cycle
- Not use submitted data to train their models
- Comply with GDPR and applicable privacy regulations

The current list of LLM providers and their DPA status is maintained in the Third-Party Risk Management Policy.

16.4 Decommissioning Alignment

When an AI/GenAI system or model is decommissioned, all associated data artifacts (embeddings, logs, evaluation data) are purged per the procedures defined in the AI/GenAI Decommissioning Policy. Decommissioning events trigger the same

confirmation and recordkeeping obligations as standard deletion requests (Sections 10 and 15).

---

17. Related Documents

- Information Governance & Records Management Standard — Data classification, lifecycle, and records disposition
- AI/GenAI Decommissioning Policy — AI system retirement and data purging procedures
- Third-Party Risk Management Policy — LLM provider DPA status and subprocessor obligations
- Incident Response Plan — Data breach and unauthorized deletion incident procedures
- User Awareness Training Program — Annual training on data retention and deletion