

What personal-data erasure requires after model training

Akintunde Idowu

akintundeidowu.com

Abstract

Deleting a source record does not necessarily resolve an erasure request once personal data have entered an AI training pipeline. This article examines the distinctions between source-data deletion, prevention of future training, output suppression, machine unlearning and retraining. It considers erasure under the Nigeria Data Protection Act 2023, drawing on comparative regulatory guidance and technical research to explain the treatment of model artefacts, backups and suppliers. Its central argument is that meaningful erasure requires an evidenced process: tracing the data, assessing what a model can reveal, selecting an appropriate remedy and explaining accurately what has and has not been removed.

Keywords

personal-data erasure; AI training; machine unlearning; data protection; accountability

Introduction

An organisation trains a speech model using recordings contributed by thousands of people. A contributor later asks for her voice recording and personal information to be deleted.

The organisation removes her name and audio files from its active dataset. The trained model remains in production. Copies of the original dataset may still exist in a vendor's environment, a development snapshot and a backup. The model may or may not retain enough influence from the recording to reproduce a phrase, reveal membership of the training set or preserve a characteristic associated with her voice.

Has the organisation deleted her personal data?

The honest answer cannot be obtained from a single database command. It depends on where the data travelled, what transformations were performed, whether the model itself contains or exposes personal information, what the law requires in the circumstances and what the organisation promised before collecting the data.

This is why “we will delete your data” becomes a more difficult statement once personal data have entered a model-development pipeline. Deletion from storage, removal from future training, suppression of outputs, machine unlearning and full retraining are different actions. They should not be described as though they produce the same result.

An erasure request concerns processing, not merely a file

Section 34 of the Nigeria Data Protection Act 2023 gives a data subject rights that include erasure of personal data without undue delay. It requires a controller to erase personal data where the information is no longer necessary for the purpose for which it was collected or processed, or where the controller has no other lawful basis to retain it.

Sections 35 and 36 add related but distinct rights. A person may withdraw consent, and a controller must make withdrawal as easy as giving consent. A person may also object to processing, after which the controller must discontinue unless it demonstrates a public interest or other legitimate ground overriding

the person's fundamental rights, freedoms and interests.

These provisions do not say that every request requires the immediate destruction of every system connected with the data. Erasure is not absolute where information remains necessary and a lawful basis supports retention. Restriction rather than deletion may also be relevant while a request, objection or legal claim is being resolved.

But the organisation cannot answer an erasure request simply by pointing to the expense of changing a model. It must identify the personal data, purposes and lawful bases that remain. It must also understand the systems in which the data are still being processed.

In a conventional records system, that may mean locating the customer profile, communications, backups and processors. In an AI pipeline, it can also require examining cleaned datasets, tokenised or transformed copies, embeddings, feature stores, checkpoints, fine-tuning files, evaluation sets, model weights, retrieval indexes, prompts, logs and outputs. The legal question follows the data through the architecture.

Deleting the training record is necessary but may not be sufficient

The first response should ordinarily be to prevent continued use of the person's source record where erasure is required. That includes the active training corpus and other identifiable copies held for the same purpose. The organisation should also ensure that the record is not silently restored during the next dataset refresh.

This step is easier where the data have reliable provenance. A stable record identifier can connect the original submission to cleaned copies, annotations, transformations and dataset versions. Without that lineage, a team may delete the obvious file while leaving duplicates in a different bucket or a supplier's workspace.

De-identification may reduce risk but does not automatically satisfy erasure. Removing a name from a distinctive voice recording does not necessarily make the recording anonymous. Pseudonymised data remain personal data where the person can be identified using additional information reasonably available to the controller or another party.

The more difficult question is what happened during training. Most models do not function as ordinary databases from which a row can simply be removed. Training adjusts parameters in response to patterns across many examples. The influence of one record may be diffuse, negligible, memorised, or recoverable only through a specialised attack. The answer depends on the model, data, training method and deployment.

That technical complexity does not make the legal right meaningless. It changes the evidence required to determine what remains.

A model is not automatically anonymous because it contains parameters

An organisation may argue that model weights are mathematical parameters rather than personal data. Sometimes that may be correct. It cannot be assumed from the form alone.

The European Data Protection Board's Opinion 28/2024 provides a useful comparative approach. It states that whether an AI model is anonymous should be assessed case by case. For the model to be treated as anonymous, it should be very unlikely both that people whose data were used to develop it can be identified and that their personal data can be extracted through queries.

The Opinion is an interpretation of the EU General Data Protection Regulation, not Nigerian law. Its reasoning is still valuable because it separates the appearance of a model from what the model permits. A file consisting of weights may continue to process personal data if information about identifiable people

can reasonably be recovered or inferred from it.

Relevant evidence may include memorisation testing, extraction attempts, membership-inference testing, access controls, the nature and uniqueness of the training data, model size and overfitting, interfaces available to users and whether safeguards can be bypassed. A model accessible only through a tightly controlled task-specific interface may present a different identification and extraction risk from the same model released with unrestricted query access.

The conclusion should also be revisited. New extraction techniques, additional datasets or expanded access can change whether identification is reasonably likely. An anonymity assessment is not a permanent label attached at the end of training.

Retraining provides a clearer counterfactual, at a cost

The most direct way to remove the influence of specified training records is to construct an approved dataset without those records and train a replacement model from that dataset. The new model provides a practical counterfactual: a version created without the information to be erased.

Even this requires careful execution. The record must be absent from every relevant source and duplicate. Derived labels or annotations may still reveal it. A base model supplied by another organisation may have encountered the same information independently. Randomness means the retrained model will not be identical to a hypothetical historical run, and testing is still needed to determine whether the output risk has been addressed.

For large models, full retraining can require substantial compute, time and engineering effort. Frequent individual requests can make immediate retraining operationally difficult and environmentally costly. Those facts are relevant to system design and proportional response. They should not be converted into a general promise that erasure stops at the dataset whenever retraining is inconvenient.

An organisation that knows before collection that its architecture cannot accommodate foreseeable rights requests should address that limitation in the data protection impact assessment, purpose and lawful-basis analysis, procurement decision and public explanation. Technical difficulty discovered after deployment is different from impossibility that was considered and justified before the processing began.

The ICO's AI and data-protection guidance makes a helpful distinction in the UK context. Erasing an individual's data from a training dataset does not necessarily require deletion of every model trained on that dataset, unless the models themselves contain the data or allow it to be inferred. Where a model contains personal data by design or leaks it accidentally, rectification or erasure may require retraining or deletion of the model.

Again, this is not binding Nigerian guidance. It illustrates the factual inquiry Nigerian controllers also need: after deleting the source records, is the model itself still processing the person's data?

Machine unlearning is a technique, not a legal conclusion

Machine unlearning seeks to remove the influence of specified training data from a trained model without always rebuilding the model from the beginning. NIST describes it as selectively removing the influence of particular training points, including to respond to a request to remove a person's records. Approximate unlearning may modify an existing model more efficiently than full retraining.

This can be valuable where a small set of records must be removed from a large and expensive model. It also creates a verification problem. A model may stop producing an obvious phrase while still retaining other information connected with the record. Behavioural tests may show that common prompts no longer

recover the data, but an adversarial query may succeed. Parameter changes may demonstrate that the model was modified without proving that the relevant influence was removed to the required degree.

The research remains active. Recent work distinguishes exact from approximate unlearning and examines behavioural and parameter-based verification. It also identifies unresolved questions about scalability, model utility, adversarial recovery and how an affected person can audit a process performed inside an organisation's infrastructure.

The legal team should therefore resist using “unlearned” as a self-proving status. The organisation needs a defined acceptance standard: what information must become unrecoverable, against which attacks, with what confidence, while preserving which legitimate capabilities? It should record the method, model version, forget set, tests, limitations and approval decision.

Unlearning may be an appropriate technical measure. Whether it satisfies an erasure obligation depends on the personal data, risk, legal basis, state of the technology and evidence that the processing has actually ceased to the required extent.

Output suppression solves a different problem

Another response is to prevent a system from returning specified information. A filter may block a person's name, a retrieval index may exclude their documents, or a system instruction may direct the model not to reproduce certain material.

These measures can reduce immediate harm, particularly while a fuller remedy is being implemented. They do not necessarily remove the information or its influence from the underlying model. A filter can be bypassed, miss paraphrases or fail when the model is accessed through another interface. Suppression is therefore closer to restricting use or disclosure than to proving erasure.

That does not make it worthless. The Nigeria Data Protection Act recognises restriction of processing in specified circumstances, and layered controls may be necessary where deletion requires investigation or a scheduled model replacement. The organisation should describe the measure accurately: “we have blocked the information from retrieval while we assess the model” is different from “the model has forgotten you”.

The distinction is especially important in communications with the person who made the request. A temporary risk-control measure should not be presented as completed erasure merely because the visible output has changed.

Backups and archives need their own deletion logic

Organisations often explain that deleted data remain in backups until the backup expires. That may be reasonable where backups are isolated, access is restricted and the retention period is justified. It becomes problematic where backups are treated as ordinary data sources, retained indefinitely or used to recreate a deleted training corpus.

The deletion process should mark the record so that restoration does not reintroduce it into active processing. If a backup is restored after an incident, the erasure instruction should be replayed before the data are used. Backup schedules, legal holds and disaster-recovery processes need to be part of the response rather than an exception nobody owns.

The same applies to model artefacts. Old checkpoints, adapters and evaluation versions may continue to contain or expose information even after the production model is replaced. Retaining them for reproducibility, security investigation or legal claims may be justified in a particular case, but the purpose, access and duration must be defined. An archive is still processing where personal data remain within it.

The request may reveal a supply-chain problem

An organisation may not control the full training pipeline. A data collector supplies records to a dataset company, which provides them to a model developer, whose model is fine-tuned by an application provider. A person may submit an erasure request to the organisation they recognise even though several entities hold related data and make separate decisions about its use.

Calling every participant a processor will not resolve that chain. Roles depend on who determines the purposes and means of the relevant processing. A dataset supplier or model provider may be a separate controller for some activities and a processor for others. The response must follow the actual allocation, not the label selected for convenience in the contract.

Section 29 of the Nigeria Data Protection Act requires a processor to act on a controller's instructions and assist with the controller's obligations, including responding to requests from data subjects. The contract should make that assistance operational.

Before training begins, the parties should agree how records will be identified across dataset versions; who receives and verifies requests; how deletion instructions move downstream; which systems and artefacts are covered; when a model must be retrained, unlearned, suppressed or retired; what evidence demonstrates completion; how subprocessors are notified; who bears exceptional remediation costs; and what happens to data and models when the relationship ends.

A generic obligation to "comply with data-protection law" does not answer those questions. Nor should the vendor be permitted to respond that deletion is impossible without explaining the architecture, alternatives and risk controls on which that conclusion rests.

The organisation communicating with the individual also needs enough information to give an accurate response. A controller cannot make a meaningful erasure promise if its supplier will disclose only that the request was "processed" without identifying what was deleted and what remains.

Privacy notices should promise a process the architecture can support

Many privacy notices say that a person may request deletion and that the organisation will erase personal data within a stated period. In an AI context, that wording may be read as a promise that the person's information will disappear from every model and output.

The notice should not be written to preserve maximum discretion at the expense of meaning. It should distinguish the main stages in terms a person can understand. The organisation can explain that it will delete identifiable source records where the legal conditions are met; prevent them from being used in future training; send deletion instructions to relevant providers; assess whether deployed models contain or permit extraction of the data; and apply an appropriate remedy, which may include suppression, unlearning, retraining or retirement.

It should also explain material qualifications: lawful retention, legal claims, secured backups, verification periods and cases where the organisation cannot identify a record without additional information. Those qualifications should not become a blanket statement that models are technically incapable of deletion.

The most dangerous promise is absolute language unsupported by evidence: "Your information is permanently removed from all AI systems immediately." A statement this broad may be technically false on the day it is published.

At the other extreme, "data used to train models can never be deleted" is also too broad. Source data can be removed, future use can be stopped, retrieval can be blocked, models can be retrained and unlearning

techniques may be appropriate. The available remedy depends on the system and the risk.

Trust comes from describing the actual process and improving the architecture so that the description remains meaningful.

Erasure readiness begins before collection

An organisation cannot design an effective deletion process after losing track of where the data went.

Before training, it should establish provenance and record identifiers; separate datasets by purpose; control duplication; define retention for raw and derived data; version datasets and models; maintain a register connecting model artefacts to their sources; test memorisation and extraction risk; and decide which model designs permit efficient retraining or unlearning.

The data protection impact assessment should consider foreseeable erasure, withdrawal and objection requests. It should identify the person authorised to decide whether deletion of the source data is sufficient and when the model must also change. That decision should combine legal, privacy, engineering and product evidence rather than asking any one discipline to declare the problem solved.

The organisation also needs a completion record. It should be able to show the request received, identity checks proportionate to the risk, systems searched, data located, legal bases considered, actions taken by each party, model assessment, residual limitations and communication sent to the person. Evidence of completion should not itself retain more personal data than necessary.

Meaningful deletion is an accountable result

Personal data do not lose legal significance merely because training has converted examples into statistical parameters. Nor does every use of a person's record mean that the resulting model permanently contains identifiable information about them.

The difficult but necessary work lies between those positions. The organisation must know what it collected, where it propagated, what the model can reveal, which lawful purposes remain and which technical intervention can be verified.

Deleting the original file may be the correct first step. It is not a universal answer. Retraining provides a stronger counterfactual but may be costly. Unlearning can reduce that cost but requires evidence of effective forgetting. Suppression can reduce exposure while leaving the underlying influence intact. Technical infeasibility may shape timing and remedy, but it should not become an untested exemption.

An organisation can meaningfully delete personal data after model training only if "deletion" refers to an evidenced process rather than a convenient interface message: stopping unauthorised use, removing records across the data chain, assessing the model itself, applying the proportionate technical remedy and telling the person accurately what has and has not been erased.

Primary and technical materials

Nigeria Data Protection Act 2023, particularly sections 24, 28, 29 and 34-36: Open source (https://ndpc.gov.ng/wp-content/uploads/2024/03/Nigeria_Data_Protection_Act_2023.pdf)

Nigeria Data Protection Commission, General Application and Implementation Directive 2025: Open source (<https://ndpc.gov.ng/wp-content/uploads/2025/07/NDP-ACT-GAID-2025-MARCH-20TH.pdf>)

European Data Protection Board, Opinion 28/2024 on personal-data processing in the context of AI models, 18 December 2024: Open source

(https://www.edpb.europa.eu/documents/opinion-of-the-board-art-64/opinion-282024-on-certain-data-protection-aspects-related-to_en)

UK Information Commissioner's Office, "How do we ensure individual rights in our AI systems?": Open source

(<https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/guidance-on-ai-and-data-protection/how-do-we-ensure-individual-rights-in-our-ai-systems/>)

NIST, definition of machine unlearning in NIST AI 100-2e2025: Open source

(https://csrc.nist.gov/glossary/term/machine_unlearning)

Google Research, "Announcing the first Machine Unlearning Challenge", 29 June 2023: Open source

(<https://research.google/blog/announcing-the-first-machine-unlearning-challenge/>)

Google Research, "New framework for auditing machine unlearning", June 2026: Open source

(<https://research.google/blog/new-framework-for-auditing-machine-unlearning/>)

Xue et al., "Towards Reliable Forgetting: A Survey on Machine Unlearning Verification, Challenges, and Future Directions", 2025: Open source (<https://arxiv.org/abs/2506.15115>)

The EDPB and ICO materials are comparative European and UK regulatory sources and are not presented as Nigerian law. Technical research is used to explain current capabilities and limitations, not to determine the legal meaning of erasure.