

THE COMPLEXITY OF ERASURE IN GENOMICA¹: ASSESSING THE IMPLICATIONS OF UTILISATION OF GENOMIC DATA FOR AI SYSTEMS & THE RIGHT TO ERASURE UNDER INDIAN LAW

Adv. Vanshika Jain¹, Dr. Aayushi Goel², Adv. Abhishek Verma³, Adv. Vinay Pathak⁴

¹Junior Research Fellow, Symbiosis International (Deemed) University, Pune. Email- vanshikajain423@gmail.com

²Assistant Professor, Symbiosis Law School, Noida, SIU. Email- aayushi.goel@symlaw.edu.in

³Bar council of Delhi. Email- adv.abhishekverma26@gmail.com

⁴Bar council of Delhi. Email- vinaypathak916@gmail.com

ABSTRACT

The growing use of genomic data in artificial intelligence (AI) systems poses complex data protection challenges, particularly regarding the right to erasure. Genomic data is inherently sensitive and may reveal details about individuals and their biological families. Removing the original dataset alone may not prevent information derived from it or inferred through the trained model when such data is used to train AI models. This raises issues regarding the practical enforceability of the right to erasure, the effectiveness of machine unlearning, and the accountability of stakeholders handling genomic data. This study assesses the effectiveness of India's data protection and health-AI governance frameworks in enforcing the right to erasure after genomic data has been used for AI training. The paper examines the alignment between statutory data-protection requirements and specialized health-AI governance guidance through a doctrinal and analytical review of the Digital Personal Data Protection Act, 2023, the Indian Council of Medical Research's Ethical Guidelines for Application of Artificial Intelligence in Biomedical Research and Healthcare (2023), and the Strategy for Artificial Intelligence in Healthcare for India. It also examines the difference between source-data deletion and model-level erasure, as well as the difficulties in determining whether technical interventions sufficiently address lingering privacy hazards. Conclusively, the paper recommends a framework that aligns the Right to erasure with the technical As a result, the research recommends a six-stage evaluation framework that is rooted in Individual Autonomy and can aid in balancing innovation and Privacy.

KEYWORDS: Genomic data; Artificial Intelligence; Right to erasure; machine unlearning; data protection; DPDP Act, 2023; India.

INTRODUCTION

"Genetic information, which can be processed from human bodily samples, contains three major types of information: first, individual traits such as the skin colour, race, eye colour, and height of the data subject; second, medical information such as information about genetic diseases and genetic predispositions; and third and most importantly, it can provide information about a class of persons who are genetically related to the data subject. Genetic data has huge application potential, from criminal investigations to assisted reproductive techniques. Thus, it is the most sensitive form of personal data." (Chandran, 2024). The integration of artificial intelligence (AI) in genomic healthcare and research is changing how genetic data is examined, interpreted, and put into practice. While artificial intelligence (AI) and machine learning (ML) are still in their infancy in genomics, researchers have already created tools that show their promise in a variety of genomic analysis and healthcare fields. By recognizing distinctive facial traits, AI-driven face analysis software can aid in the diagnosis of genetic diseases. Moreover, machine learning approaches have been used to forecast the course of cancer in individual patients and determine the major site of cancer using data from liquid biopsy. Additionally, ML-based methods can aid in distinguishing pathogenic genomic variations from benign ones, therefore aiding in the interpretation of genetic information. In addition to CRISPR, deep learning has been employed to increase the accuracy and effectiveness of gene-editing techniques. In addition, these applications show that AI may help with the identification, interpretation, and use of genomic data. (National Human Genome Institute 2022). "Despite these unprecedented opportunities, serious concerns have surfaced regarding genomic data privacy. Unlike most other forms of medical information, genomic data possesses a unique and highly sensitive nature. It not only reveals the genetic identity of an individual but may also disclose information about their relatives and even future generations. Consequently, any misuse or leakage of genomic information can have far-reaching implications that extend well beyond the individual". (Kohan, 2026). A fundamental principle of data protection legislation is that personal data may be deleted in appropriate situations if the purpose of its collection has been achieved, if consent is revoked, or if another legally recognized cause for deletion occurs. And if then "even anonymized genomic data carries the risk of re-identification, particularly when combined with other datasets such as clinical or digital health records." (Kohan, 2026). The use of personal data to train an AI model reduces the certainty of its applicability. During the training process, data affects the patterns a system learns and the tuning of model parameters. Therefore, deleting the initial entry does not automatically prove that the information no longer affects the model or that the model cannot reveal data related to the training data. (Longpre et. al., 2023). Even if the original data is

no longer in the training set, a system trained on genomic data may still draw inferences. Therefore, the legal concern is not merely whether a record has been erased. When a model's learned behavior allows information to persist, it raises questions about what deletion means. In April 2026, a Stanford study titled "Why AI Cannot Forget Genomic Data" highlights this strain. (Sutthivana, 2026). The study assessed the difficulty in enforcing the right to be forgotten in AI systems trained specifically on genomic data, the challenges of machine unlearning, and the implications of genetic information for biological relatives. The analysis begs a broader question concerning the interaction between artificial intelligence systems whose acquired patterns may continue to be valuable for research while also posing privacy concerns and individual data protection rights especially when Data protection regimes across the globe keep genomic data under "sensitive data" or require them to be de-linked from AI systems once their collection purpose is fulfilled. For the purpose of this study, the author studies the "Right to Erasure" in the context of the specific Stanford study with India's data privacy law, i.e., the Digital Personal Data Protection Act (DPDP Act). Subject to the requirements outlined in the Act, Section 12 acknowledges the right of a Data Principal to request the modification, completion, updating, and deletion of personal data. Where the Data Principal withdraws consent, or the specified purpose is no longer being fulfilled, Section 8(7) mandates erasure, subject to the legal prerequisites and exemptions. However, more thorough research is required to determine how such provisions apply to information included in a trained AI model. The nature of AI training, the legal definition of personal data, the requirements for deletion, and the difference between deleting source records and handling information that might still be recovered from a model must all be considered when interpreting the statutory language. Additionally, the question necessitates considering the legal standing and range of India's healthcare AI governance frameworks. The "Ethical Guidelines for Application of Artificial Intelligence in Biomedical Research and Healthcare" (2023) from the Indian Council of Medical Research also deals with issues such as consent, privacy, data minimisation, secondary use, and data deletion or modification while using AI for healthcare, while the "Strategy for Artificial Intelligence in Healthcare for India" (SAHI, 2026) takes a lifecycle-based approach to healthcare AI governance and addresses data governance, privacy safeguards, representativeness, and continuous oversight. These guidelines are pertinent to the governance of genomic AI; however, it is important to separate their unique legal force and the degree to which they address model-level erasure from the binding duties established by law.

This study therefore asks: *what is the extent to which India's data protection and healthcare AI governance framework addresses the right to erasure where genomic data has been integrated into an AI model, and what additional legal and technical safeguards may be necessary?*

The research is based on the assumption that the presence of a statutory right to erasure does not, by itself, respond to the technical complications posed by AI systems and Machine unlearning. In furtherance of such issues, the present paper examines whether the present privacy frameworks in India sufficiently distinguish between the removal of source data and the retention of information through trained models, whether machine unlearning can lead to significant compliance, and how erasure may be checked when models are reused, changed, or included in downstream systems in context to genomic data where the impact can go beyond the individual, and can ascertain and reveal information regarding biological relatives, and shall try to also balance it with Right to erasure and how the far aching of such right may hinder legitimate research and healthcare activities.

The paper takes a doctrinal approach, examining the DPDP Act, the rules and commencement clauses that apply, pertinent Indian healthcare and AI governance tools, and chosen comparative legal and technical literature. It considers whether current principles may address the particular difficulties related to genomic data and model training, rather than assuming that genomic AI necessarily necessitates a separate legal framework. The goal is to determine the exact legal and governance gaps, if any, and to suggest safeguards that reconcile legitimate scientific and healthcare goals with significant data protection.

Genomic Data In Ai Systems- From Data Processing To Model Retention

"Artificial intelligence (AI) and machine learning (ML), especially deep learning, have profoundly transformed biology by enabling precise interpretation of complex genomic and proteomic data." (Hein et al., 2025). Information about DNA sequences, genetic variations, and other molecular characteristics that may aid in the identification of disease-related mutations, the comprehension of biological processes, and the creation of customized treatment programs makes up genomic data. Utilisation of AI and ML techniques aids and "offers significant potential in genetic engineering and therapeutic gene therapy, helping to create new hypotheses and experimental techniques. It can detect heredity and gene-related disorders from gene models and is crucial for rational drug discovery and design." (Dhaarani & Reddy, 2025).

It can also use machine learning to detect patterns in genomic datasets, categorize genetic variations, and predict outcomes that can help with additional scientific research and clinical interpretation. "One of the most transformative applications of AI in genomics lies in personalized medicine, where treatments are tailored to an individual's unique genetic makeup (Schork, 2019). By leveraging AI-driven insights from genomic data, clinicians can predict disease risk, select optimal therapies, and monitor treatment responses more effectively than ever before (Ozaki et al., 2024). For example, in oncology, AI models can analyze tumor genomes to identify actionable mutations and recommend targeted therapies, significantly improving patient outcomes (Ozaki et al., 2024)." (Dara et al., 2025). In a similar vein, O'Connor and McVeigh (2025) underline the increasing significance of AI in genomic medicine, especially in cancer care, where the examination of extensive genomic datasets may facilitate earlier cancer diagnosis, treatment selection, and prognostic evaluation.

These technologies have practical uses in several phases of genomic analysis. AI-powered facial analysis may help identify genetic illnesses by detecting distinguishing phenotypic characteristics, while machine learning models have been created to predict cancer progression and determine the most probable main site of cancer using liquid biopsy data. Computational models can also aid in differentiating between pathogenic and benign genetic variations, helping in the interpretation of

genomic results. Furthermore, deep learning methods have been used in CRISPR and other gene-editing technologies, as well as in the forecasting of the functional effects of genetic variation (Mahmud & Banerjee, 2026; O'Connor & McVeigh, 2025). (Dara et al., 2025). AI is being integrated into genomic procedures including sequence analysis, variant calling, annotation, and interpretation, showing that its function is not limited to a single diagnostic assignment (Illumina, n.d.). These advances highlight the promise of AI in turning genomic data into insights that might help healthcare and research choices. However, the expediency of these applications is greatly impacted by the dependency of these models on diverse representations in the data, the quality of data, and, most importantly, the availability of such data. The increased usage of genomic datasets to build AI systems for deployment in several clinical and research environments raises concerns about the purposes for which such data may be processed, how long it may be kept, and how much information about people remains protected after model training.

From Model Training to Genomic Information Preservation

During training, machine-learning systems adjust model parameters to recognize patterns in the training data. The resulting model is not necessarily a replica of the dataset it was trained on. However, information from training samples may be recalled or accessible through certain types of interaction, depending on the architecture, training procedure, dataset features, and subsequent access to the model. In other cases, a model may retain statistical correlations discovered during training without creating a recognizable record. These options need to be differentiated: the influence of a training record on model behavior does not, in and of itself, prove that the model stores or releases a person's genomic information. A Harvard study also reveals that "In a real-world example, a model might be trained to avoid generating certain types of content, but this doesn't necessarily mean the underlying information has been removed from its knowledge base." (HBS AI Content & Learning et al., 2025). Several mechanisms by which information related to training data may be compromised have been studied in research on machine learning privacy. Data-reconstruction and model-inversion approaches attempt to infer or recover information about training examples, whereas membership-inference assaults attempt to ascertain if a specific record was part of a model's training data. These attacks' efficacy varies according to the model, the data, the attacker's expertise, and the amount of access offered. They highlight why the privacy implications of training cannot be assessed simply by examining the storage of the original dataset, rather than suggesting that every trained model is susceptible to successful reconstruction. Chourasia & Shah (2023) further show in their examination of machine unlearning that official assertions regarding the eradication of a record's influence must take into account the privacy guarantees offered by the deletion procedure itself.

These issues are especially pertinent to genomic applications. Rare variants or feature combinations in genomic datasets might heighten the sensitivity of any information that is revealed or inferred from them. Although an AI model may not duplicate a whole genomic sequence, its results may reveal information relevant to an individual or support assumptions about that person in certain circumstances. It should not be presumed that the risk will always exist and that it is dependent on the characteristics of the model and the data. However, since genomic privacy cannot be evaluated simply by asking whether a full sequence can be recovered from a trained system, the possibility of inference is significant. Hence, it is crucial to make a distinction between a trained model and a source dataset. When a genomic record is removed from a database, it ensures that it will no longer be available in that database. It does not automatically prove that the record's contribution to the model has been removed or that a previously trained model no longer retains information connected with the record. Additionally, the fact that a model was trained on a certain record does not mean that the record is still recoverable or that the model needs to be retrained. The issue is whether the deletion procedure in this system can be shown to result in a result and what it does. This differentiation becomes more significant when genomic data goes through several stages of processing. A dataset may be utilized to develop a model that is subsequently fine-tuned for another reason, or it may be cleaned, converted, or merged with other datasets. Models can also be included into downstream applications or made available to other academics. Therefore, a request to erase the original record may need to take into account additional copies, derived datasets, or models that were trained using the pertinent information, depending on the system's design and governance structures. Borkar et al. (2025) illustrate that "unlearning can cause additional privacy risks: can lead to stronger membership inference attacks and that unlearning can increase membership inference accuracy on other training samples." The existence and scope of these dependencies must be established for the system concerned; they should not be assumed merely because the original data was employed in training.

Thus, the technical issue is not limited to whether genomic data may be removed from a storage place. The question is whether the applicable information may be found during the data-processing lifecycle, whether a suitable approach can address its existence or impact in a trained model, and whether the efficacy of that method can be evaluated.

The Limitations of Machine Unlearning and Technical Erasure

"Machine learning (ML) seeks to automatically learn meaningful relationships and patterns from examples and observations (Bishop 2006). Advances in ML have enabled the recent rise of intelligent systems with human-like cognitive capacity that penetrate our business and personal life and shape the networked interactions on electronic markets in every conceivable way, with companies augmenting decision-making for productivity, engagement, and employee retention (Shrestha et al. 2021), trainable assistant systems adapting to individual user preferences (Fischer et al. 2020), and trading agents shaking traditional finance trading markets (Jayanth Balaji et al. 2018)". (Janiesch et al., 2021). Its growth has been fueled by the high computational cost of rebuilding models from scratch whenever a record has to be removed. Depending on the model and method, unlearning may involve changing model parameters, eliminating the effect of certain training examples through a specific technique, or retraining the model on an updated dataset. These approaches vary in terms of the assurances they can offer and their computational needs. The effectiveness of machine unlearning is still a major technological issue. One must make a distinction between trying to delete a record from a training dataset, deleting it, and

demonstrating that the deletion was successful. Data administration, model-level intervention, and validation are the subjects of the first, second, and third points, respectively. Completing one action does not mean that the others have also been completed, even if they are part of the same erasure procedure. (Xu, 2018). Before assuming that the similarity between an unlearned and a retrained model, or the application of an unlearning method, is reliable evidence that all relevant privacy concerns have been addressed. In addition, Chourasia and Shah (2023) show that retraining alone cannot always resolve all privacy concerns associated with deletion because information may be inferred from relationships between model releases and data. Alternative tactics try to calculate the consequences of retraining by making certain adjustments to an outdated model. Although such strategies may reduce processing costs, the amount of essential data erased depends on the method's presuppositions, application, and assessment. When studying genomic data, these limits are essential. A model trained using genomic data can be adapted or integrated into other systems for tasks other than the one for which the original data was gathered. Any assessment of an eradication procedure must be made in consideration of the particular model and the applicable data-processing procedures at the time it occurs. (Davis-Turak et al., 2017). It cannot be assumed that removing a record from a dataset will address every model and derivative system linked to that dataset. Similarly, one should not presume that a person's removal request automatically leads to the elimination or retraining of every derived model. The right technical response is determined by the system architecture, the information at hand, the viability of the available solutions, and the necessary level of assurance.

Another difficulty arises in verifying whether a machine has forgotten anything. Verification aims to establish if a model has successfully removed the information or impact that was sought. Current methods may be broadly categorized based on the data they take into account. In order to ascertain whether model outputs or replies continue to give information associated with data that has to be erased, behavioral methods examine these outputs or responses. Parametric methods are utilized to analyze internal data pertaining to model parameters or models. It is crucial to remember that each approach has its drawbacks, and a single test outcome should not be regarded as absolute evidence that all data has been completely erased. However, a model's inability to reveal a record under a certain test does not necessarily demonstrate that no valuable information can be gleaned under different circumstances. As a result, the unlearning technique itself is independent of the authentication challenge. Even though a vendor might delete or unlearn a process, there are still two separate problems: the action's completion and its efficacy. Being particular about what needs to be removed, what model or data-processing components are involved, what effectiveness standard is being employed, and what evidence supports the discovery is essential for performing a meaningful evaluation. Without a universally applicable verification method, it's particularly important to distinguish between a confirmed result and a documented technology intervention (Qureshi et al., 2025). Ongoing concerns regarding the correctness, consistency, and limitations of existing verification procedures are brought to light by research on unlearning verification. Due to the potentially delicate nature, difficulty in replacement, and relevance to both the person involved and their biological family, these issues are of even greater importance for genomic AI.

As a result, the scientific research does not establish that removing elements at the model level is consistently achievable nor that it is inherently unachievable. Instead, it highlights a problem with method, scope, and assurance: the available measures vary across systems, and the sufficiency of a deletion claim depends on what was removed and how that result was evaluated. It provides the technical foundation for evaluating whether the Indian data protection frameworks and health-AI governance adequately handle erasure requests once genomic data has been used to train an AI model.

Privacy Ramifications of Persistent Genomic Information

The privacy implications of genomic AI extend beyond the collection and storage of genetic information. "Human genome sequences are often considered sensitive data as they can reveal personal information about their owners; with access to a genome and publicly available Genome-wide association study (GWAS) summary statistics, it is possible to infer an individual's ancestry, his relatedness to others and their risk of developing certain diseases. The misuse of such inferred information may result in 30 genetic discrimination." (Longpre et al., 2023). These issues are especially pertinent when a subsequent request to delete the underlying data is made, and genomic information has helped train an AI model. The value of such a request depends on the accessibility of the original record, the sort of information that can be deduced, the individuals who may be impacted, and the ways in which the resulting model is utilized. "The re-identification of a genome creates a privacy risk because of the sensitive information a genome 286 could disclose about his owner" (Cavinato et al., 2026). Despite differences in their likelihood and severity depending on the characteristics of the dataset and the environment in which it is handled, the privacy risks associated with genomic data are well-known. In a scoping study of proven privacy threats on genetic datasets, Thomas et al. (2024) identify identity tracking, inference, and membership attacks as distinct categories of risk. According to their findings, a single metric, such as whether direct identifiers have been eliminated, cannot be used to determine genomic privacy. Instead, the evaluation should take into account the content of data and data structure, the availability of outside information, and the practical viability of connecting or inferring information from the dataset. These issues become particularly crucial when AI systems that may be shared or accessed outside of the organization that first gathered the data handle genomic data.

Re-identification, Inference and the Limits of Anonymisation

The potential for re-identification is a major privacy concern in genomic data processing. Simply eliminating names, contact details, and other direct identifiers does not guarantee the anonymity of genomic data. To identify an individual, genetic traits can be coupled with outside datasets or genealogical information; meanwhile, a combination of demographic and genomic data can raise the likelihood of a successful identification. Thomas et al. (2024) show that the risk is determined by the distinct features of the genetic data as well as the resources accessible to someone trying to identify or deduce information about a data subject. Therefore, genomic data shouldn't be considered uniformly identifiable in all situations, and the removal of direct identifiers shouldn't be taken as definitive proof of anonymity. "Moreover, even the

sharing of de-identified data for medical research typically faces tension between open sharing within the research community and exposure to privacy risks. These risks generally stem from the ability of some data recipients to link the genomic data to the identities of the corresponding individuals. One particularly acute concern raised in recent literature is the ability to link a genome to the photograph of an individual's face. Specifically, these studies have shown that one can effectively match high-quality three-dimensional (3D) face maps of individuals with their associated low-noise sequencing data, leveraging known associations between phenotypes, such as eye color, and genotypes, which, for the purposes of this study, correspond to the variations in our genes that affect physical traits." (Venkatesaramani et al., 2021) AI introduces an additional dimension to this concern because trained models may support inferences about their training data. Chen et al. (2021) examine membership-inference attacks against machine-learning models trained on genomic data. Such attacks attempt to determine whether a particular record was included in a model's training dataset. In the genomic context, evidence of participation in a study may itself reveal sensitive information, particularly where the study concerns a specific disease or health condition. "Furthermore, advancements in machine learning and Bayesian networks have introduced novel threats to genomic data security. Bayesian networks have been used to predict missing genomic markers and perform identity inference. This remains possible even when differential privacy protections are applied" (Annan et al., 2025). Bonomi et al. (2022) found that models may reveal data about the participation of individual data contributors. This is illustrated by the membership-privacy risks in cancer genomics. The implications extend beyond identifying a person. Inferences about disease susceptibility, inherited characteristics, and biological relationships may be made using genomic data. The reliability of the results and the amount of data at hand dictate how susceptible such inferences are (Royal et al., 2010).

Even if an inferred quality is ambiguous or erroneous, the inference itself can still pose a privacy risk if it is treated as established information or used to guide significant choices. Consequently, it is essential to draw a distinction between revealing an already-existing record and generating fresh data about an individual by analyzing genomic data.

Re-identification also makes it difficult to determine whether data has been removed. The original deletion may not address all risks that come from the wider processing environment if an associated model, a downstream system, or another dataset still allows access to information that is related to a record that has been removed from a source dataset (Simon et al., 2019). This does not mean that re-identification will necessarily remain possible after deletion. Rather, it indicates that the effectiveness of an erasure measure must be considered in relation to the specific information and systems involved. The relevant concern is whether the remaining information can still be linked to an identifiable person or used to infer sensitive attributes, and whether the measures taken are adequate to address that risk.

For genomic AI, therefore, the distinction between anonymisation and erasure is important. Anonymisation seeks to reduce the possibility of identifying an individual from information that continues to be processed or retained. Erasure, by contrast, concerns the removal of data within the relevant scope of a request or obligation. Neither concept should be assumed to guarantee the other. Removing identifiers does not necessarily eliminate the possibility of re-identification, while deleting a particular record does not, without further assessment, establish that every related privacy risk has disappeared. This distinction provides a necessary foundation for evaluating the consequences of continued processing of genomic information.

Familial Privacy and the Relational Nature of Genomic Information

"Genetic data are not limited to one individual, with information about one person revealing information about the person's close and distant biological relatives. Only by studying genetic information from many people can the significance of the individual's variants be discerned. The importance of understanding the causes of health and disease has led some to argue that people have some obligation to share data about themselves for low-risk research. This public nature and value of the genome makes it difficult to decide what level of control individuals should have and how to provide appropriate privacy protections" (Clayton et al., 2019). The privacy implications of processing genomic information are consequently not always confined to the individual whose record is held by a researcher, healthcare provider, or AI developer. Such identification can affect People who have not participated in the initial data collection or agreed to the subsequent use of their family members' information. The resulting concern is relational: even if the individuals have no direct connection to the organization processing the genomic information, a person's decision to disclose it may have privacy implications for other family members. "It has been shown that it is foundationally insufficient to address the challenges and opportunities presented by blooming of genomic data, where one individual's decision regarding genomic information to consent to data sharing can directly impact the privacy of their biological and social relations" (Huang et al., 2026).

This relational aspect becomes increasingly crucial when genomic data is integrated into AI systems. Depending on the knowledge gleaned and the task, a model trained using one individual's data may aid in making inferences about biological traits or relationships that are shared within a family. Therefore, the implications for erasure are equally complex. A request made by one individual cannot automatically be treated as a request on behalf of every biological relative and at the same time, removing one person's information may not erase information about every related individual that can be inferred from other available records or datasets. The familial dimension therefore complicates both the scope and the assessment of privacy protection.

This problem also reveals a limitation in genomic privacy. Individual consent and deletion remain important, but they may not capture all the consequences of processing gene information which is shared across biological relationships. It is not that individual rights become irrelevant; rather, it infers processing and erasing of gene data is a part of a larger complex continuum of information in which one person's data may affect the privacy interests of others. Any regulation requirement of genomic AI must therefore consider both individual-level risks and the relational implications of processing.

Secondary Use, Loss of Control and the Consequences of Continued Processing

Genomic data is typically beneficial since it can be reused for research other than the reason for which it was first collected. Data can be used to research illness patterns, create novel diagnostic techniques, or improve predictive models. Significant scientific and healthcare benefits might come from such reusability of gene data. But it might also establish a gap between the data's original collection purpose and its subsequent uses. It may become more and more difficult for the data subject to follow their information's path or comprehend how it is still being utilized as datasets are merged, models are created, and outputs are integrated into additional systems.

The issue is not that secondary use is inherently inappropriate. Further research may be beneficial and may be conducted under suitable ethical and governance arrangements. "Data types vary in identifiability and re-identification risk, and safeguards vary in scope, underscoring the importance of examining what institutional policies specify about conditions for sharing without explicit consent while recognizing that such policies cannot supersede the broader legal frameworks within which institutions operate"(Ju et al., 2026). The concern arises when the purposes, recipients, and consequences of continued processing are not sufficiently transparent or when the arrangements for managing data no longer reflect the expectations under which the information was originally supplied. "From an ethical perspective, one of the most important problems with such reuse is the consent of the person offering the data and to what extent and exactly how the individual concerned should be informed about reuse of that data. As Petrini points out, many samples and biological data are obtained in situations in which the person has not given their consent or has not given it for future reuse" (Lorenzo et al., 2023). In AI-assisted genomic research, a record may contribute to a model whose later uses differ from the original research question. The extent of this possibility depends on the terms of access, the design of the research and the arrangements governing the model's subsequent use.

Continued processing may also make the practical consequences of a deletion request more difficult to assess. Where genomic data remains confined to a clearly defined dataset, its removal may be comparatively straightforward to document. Where the data has been incorporated into training pipelines, derived datasets or models that have been shared with other organisations, the scope of the required response may be less apparent (Guy & Boonthum-Denecke, n.d.). A record's removal from one system does not itself establish what has happened in other systems, nor does the existence of a downstream model establish that the individual's information remains recoverable from it. The relevant question is whether the data-processing chain has been sufficiently mapped to identify which components require action and how the result can be assessed.

Another concern raised in this ecosystem is an individual's capacity to meaningfully assert autonomy over their own information. Exercising the Right to erasure becomes challenging when a source data may be erased from an AI system, but even the regulator cannot essentially determine whether the request has been carried out throughout the appropriate processing environment. This is an accountability issue as much as a technical problem. Therefore, the issue is not whether deletion is computationally possible, but also whether the processing organization can explain the scope of its actions and provide a suitable foundation for evaluating their efficacy.

These concerns must be balanced against the legitimate value of genomic research. Scientific progress often depends on the responsible use of large datasets, and on taking effective measures that reduce privacy risks. Chen et al. (2021), for example, examine the trade-off between privacy protection through differential privacy and predictive accuracy in genomic machine-learning models. Their findings illustrate that privacy safeguards may involve practical trade-offs rather than offering a cost-free solution. Such trade-offs do not negate the importance of privacy protection; they demonstrate the need to evaluate safeguards according to the relevant risks, technical context and legitimate purpose of processing.

Thus, the impact of persistent genomic data may be viewed across three interrelated categories: First, continued use of genomic information may result in identification and disclosure concerns. Second, because genetic information might disclose information about biological relatives, such risks could go beyond the first data subject. Third, once information has been entered into an AI development lifecycle, the reuse and dissemination of datasets and trained models may make it difficult to exercise and verify control over it. While these categories are similar, each introduces a new issue that needs to be considered when determining the sufficiency of privacy protection. At the same time, the evidence does not support the conclusion that every trained model necessarily retains identifiable genomic information or that every instance of secondary use produces a privacy violation.

The Right to Erasure under Indian Law

When personal data has already been used to train an AI model, what happens to an individual's privacy, especially when such information is genomic information that has both direct and inferential consequences, information that does not affect the individual but their family as well? The scope of "Right to Erasure" is a fundamental question that is raised when genomic information is used to train AI systems. Typically, data-protection systems handle erasure by eliminating personal information from the systems where it is handled or kept. However, as already identified by Sutthivana (2026), AI systems are unable to effectively unlearn the genomic information; what becomes of the right to erasure, and individual autonomy then? The legal issue is not just whether a right of erasure exists, but whether the applicable framework adequately handles its execution, scope, and verification following model training.

This section of the paper addresses this issue using three supplementary aspects of India's governance framework: the 2023 Digital Personal Data Protection Act (DPDP Act), the 2023 Indian Council of Medical Research's Ethical Guidelines for Application of Artificial Intelligence in Biomedical Research and Healthcare, and the Strategy for Artificial Intelligence in Healthcare for India (SAHI), which was introduced in February 2026. These instruments are not to be regarded as legally equivalent. Subject to its commencement provisions and relevant exceptions, the DPDP Act creates statutory rights and responsibilities. SAHI offers strategic guidance for responsible AI adoption in the health sector, while the ICMR standards provide ethical guidance for AI in biomedical research and healthcare. The specific issue of the right

to erasure when genomic data has previously been utilized to train an AI model must be used to weigh their individual contributions.

The Right to Erasure under Indian Data-Protection Frameworks

“The right to be forgotten consists of two dimensions: the right to forget and the right to delete (including the right to object). The right to forget means ensuring individuals are not permanently bound to their past, thereby safeguarding human dignity by allowing the possibility of being “forgotten”. Where a data subject does not wish their personal data to be processed or stored by a data controller and there is no legitimate reason for maintaining such data, it should no longer be made accessible to the public”(Tyagi, 2026). This distinction is especially relevant to genomic AI. The source record may be held in a research database, while a trained model may continue to be used for a related or different purpose. A request to erase the record raises questions about which data and processing procedures are covered by the request and whether the applicable duties apply to information related to the trained model. Technical uncertainty does not, on its own, eliminate a legal obligation, but it also does not mean that every possible technological outcome of training must be reversed simply because there is a right of erasure. The legal standard and how it is met technically must therefore be considered separately.

The DPDP Act of 2023 establishes the primary legal framework for processing digital personal data in India. The Right to Erasure finds an explicit mention in the regulation under section 12 wherein section 12(1) provides that a data principal i.e. the owner of the data can request the processor and user of such data to erase the data in compliance with the Act. However, section 12(3) also provides the condition that the data shall be erased, unless its retention is necessary for the specified purpose or in accordance with any legislation. In addition, the act under section 8(7) also provides that a data fiduciary i.e the processor of the data shall erase personal data as soon as the data principal revoke its consent or the purpose for which the consent was asked is fulfilled, but their application to genomic AI requires careful interpretation. The statutory language addresses personal data and its processing; it does not expressly set out a dedicated procedure for removing genomic information from trained AI models nor does the act make any distinction/ category like other International regulations such as GDPR called “sensitive data”.

Further, the application of these provisions to train AI models needs a closer look. Genomic data maybe collected for healthcare or biomedical research and subsequently being utilized to develop, train or optimize AI systems. If a request to erase the gene data is submitted after training, then as already identified in sections 2 and 3, AI systems cannot effectively unlearn or delete or erase a set of data, especially genomic data. It is also important to differentiate between the original record and the trained model in this case. Deletion may be directly implemented against a genomic record if it is still present in a dataset. If the organization has already utilized the data for training, it must determine which personal data remains within the processing environment and if additional steps are required. The removal of a database entry should not be considered to have satisfied the statutory obligation if the pertinent personal data is still being processed elsewhere. Likewise, without additional technical and legal assessment, the simple reality that a model was trained on a record is not enough to conclude that the model itself contains identifiable personal data or that a particular unlearning technique is legally required.

A further consideration is the relationship between erasure and research. Section 17(2)(b) of the DPDP Act provides a conditional exemption for processing necessary for research, archiving, or statistical purposes, provided the personal data is not used to make a decision specific to a Data Principal and the processing adheres to prescribed standards. This provision is essential for genomic research because it identifies a sort of processing that may be used for a wider range of scientific objectives. But it should not be seen as a broad waiver for all commercial AI training or research. Its applicability is determined by the statutory requirements, the processing’s nature and objective, and the relevant criteria. A fact-specific assessment is necessary to determine whether a genomic AI activity is covered by this section.

The timing of implementation also matters. Although the DPDP Act was enacted in 2023 and the Digital Personal Data Protection Rules, 2025 were notified on 14 November 2025, the framework has a phased commencement schedule. As of 21 October 2026, the core provisions in sections 3–17, including section 12, are scheduled to commence eighteen months after publication of the notification, on 13 May 2027. The relevant obligations must therefore be described as enacted provisions scheduled for commencement, rather than assumed to be fully operative at the time of this analysis.

Thus, the DPDP Act creates a significant legal foundation for erasure. However, when it is applied to genomic data utilized in model training, questions are raised regarding the connection between continued processing, source-data deletion, and the evaluation of residual information. It is vital to consider these questions in light of the focused guidelines such as ICMR & SAHI 2026.

The ICMR Guidelines and Ethical Governance of AI in Biomedical Research

The Ethical Guidelines for the Application of Artificial Intelligence in Biomedical Research and Healthcare (2023) of the ICMR offer sector-specific ethical GUIDELINES for the creation, implementation, and adoption of AI in healthcare and biomedical research. The standards address ethical review, informed consent, stakeholder responsibilities, ethical principles, and governance. Their value for genomic AI stems from the fact that they understand that medical AI’s ethical use goes beyond simply technical performance; the privacy, transparency, consent, and accountability of health-related information’s collection, processing, and use must also be taken into account. The standards highlight issues with combining location, social-media, genetic, medical, and other data sets. Even when privacy-protective measures have been applied to individual datasets, there is still a chance that re-identification may occur when information from various sources is combined. Additionally, the guidelines highlight the value of responsible data sharing, suitable safeguards for personal information, and careful consent when data is utilized for purposes other than those initially considered. Since

genetic information's privacy implications might go beyond the situation in which it was initially obtained, these issues are pertinent to genomic AI.

Additionally, the guidelines offer a foundation for taking into account the administration of data throughout its lifecycle. Their discourse on data minimization, data retention, the utilization of previously acquired datasets, and the deletion or modification of data following an opt-out or right-to-be-forgotten request is relevant to the governance of AI systems that process biomedical data. Such recommendations help in ensuring that privacy concerns are addressed not only when data is collected but also when it is maintained, repurposed, or used for a different purpose than initially intended. However, it is necessary to carefully isolate the implications of these limitations for trained genomic AI models.

Directions for deleting data do not always establish how information that has already been integrated into a trained model should be handled. The regulations do not by themselves establish a thorough technical standard for machine unlearning, propose a universal approach for model-level erasure, or create a complete verification protocol to show that a model no longer maintains or reveals data related to a certain person. Ethical principles can offer advice that informs institutional practice, data governance, ethics committee review, and research design without explicitly detailing the technical implementation of each privacy protection. Therefore, their input differs from that of a statute. Whether and how the ethical requirements established using ICMR recommendations are adequately implemented in practice and via linked governance structures to handle erasure requests involving trained models. This discrepancy raises questions about the responsibilities of developers, institutions, and academics in relation to genomic AI. Responsible administration may require thorough documentation of the data's origin, the purposes for which it is used, the terms of future sharing, and the procedures available when consent is revoked, or erasure is requested when a genomic dataset is used to train a model. If a model is shared or modified later, the initial research organization may not be the only one contributing to its continuous use. Consequently, the adequacy of the accords will be determined by the governance systems that establish acceptable systems, assign responsibility, and record the procedures used in response to a request.

Thus, the ICMR suggestions provide an appropriate ethical framework for preserving genomic data, particularly with regard to informed consent, privacy safeguards, data reuse, and lifecycle administration. However, the question of implementation and technical detail continues to be brought up in relation to their connection to model-level erasure. The rules cannot be enforced, but they may be used to assess what ethical genomic data handling entails once the data has already been used for AI training.

SAHI and the Governance of Healthcare AI Across the Data Lifecycle

The Strategy for Artificial Intelligence in Healthcare for India (SAHI), which was introduced by the Ministry of Health and Family Welfare in February 2026, offers national strategic direction for the ethical use of AI in India's healthcare system. SAHI is a framework for ethical, evidence-based, inclusive, and safe adoption, according to the government, with a focus on governance, data stewardship, validation, deployment, and monitoring. It is pertinent to genomic AI applications, including those that employ sensitive genetic information for research, prediction, or clinical support, due to its emphasis on ethical healthcare AI. The ICMR standards and the DPDP Act are different from SAHI's contribution. It offers a strategic framework for managing the growth and application of healthcare AI, as opposed to a unique statutory appeal to delete personal information. As a result, it should not be taken as independently establishing a legally binding obligation to AI models for unlearning gene data. Its significance is found in the institutional frameworks and governance guidelines that can be used to monitor, assess, and create healthcare AI systems.

The privacy repercussions of data usage can vary over a system's life cycle; therefore, data stewardship is essential for genomic AI. It is possible to collect data for research, combine it into a dataset, employ it to train a model, and then process it through deployment, monitoring, or adaptation. Moreover, the model and data could be exchanged across institutional boundaries. Issues that arise following training when a deletion request is made may not be fully resolved by a governance plan that only takes into account pre-deployment assessment or initial collection. SAHI's emphasis on administration and supervision provides a useful framework for determining if responsibilities are clearly allocated throughout these stages. However, Healthcare AI governance may need to pay attention to how data sources are recorded, how model development and adaptation are monitored, how privacy risks are assessed, and how institutions respond to problems that occur after implementation.

Nevertheless, a certain erasure norm is not automatically established by the existence of lifecycle management. Watching a model's performance, determining its clinical suitability, and evaluating its durability are not the same as showing that data associated with a person has been erased. In a similar vein, without more specific procedures, broad commitments to data privacy and responsible usage do not outline how a provider should evaluate the efficacy of model-level erasure. The distinction between general governance goals and a specific erasure procedure must thus continue to be clear. The strategic importance of SAHI should not be underestimated for this reason. A national framework may help set expectations for institutional cooperation, ethical implementation, and evaluation of healthcare AI systems. Additionally, it might create an environment for creating more comprehensive sectoral regulations for sensitive information, such as genomic information. The question is if these strategic goals are reflected in practical steps that address the unique problems of post-training erasure, such as pinpointing impacted systems, allocating accountability, and demonstrating compliance.

Consequently, SAHI should be viewed as complementing the larger governance framework rather than replacing the DPDP Act or the ICMR recommendations. The three tools address various aspects of ethical genomic AI: legal rights to data protection, ethical governance in biomedical research, and strategic oversight of healthcare AI. The data processing life cycle's ability to be managed on a unified basis for handling erasure requests determines their combined applicability.

There is a lack of operational clarity and clear enforcement when these three regulatory tools are taken together. The current framework serves as a foundation for defending personal data and encouraging sensible health-AI governance but

is ineffective in regulating genomic information and providing a suitable technical solution for asserting the Right to Erasure.

Recommendations and the Way Forward: Towards a Verifiable Erasure Framework for Genomic AI in India

The preceding analysis in the paper indicates that even erasing/ deleting genomic information from a source database may not address all issues associated with its utilization in artificial intelligence training. Depending on the specific model and the type of information involved, privacy threats may remain via derived datasets, subsequent sharing, or information preserved in the trained models. Therefore, India necessitates a strategy that links legal responsibilities with technical verification, responsible management of genomic data, and accountability. The WHO (2024) recommendations regarding the collection, access, use, and sharing of human genomic data further reinforce the need for transparent consent, responsible secondary usage, and accountability. The Indian framework is currently scattered and is working through soft-law obligations for regulation of genomic information used for training of AI systems. The ICMR guidelines and the SAI 2026 both have placed emphasis on the erasure of genomic information from the AI systems however, such erasure is not effective as already been identified through this paper therefore, India ought to opt for a more effective and central regulation for management of genomic information that is guided by privacy jurisprudence and is technically viable.

For the purpose of Erasure requests from AI systems, India can look at the mechanism used by Ginart et al. (2019), which shows the importance of designing machine-learning systems that can aid in data deletion. By developing such systems, institutions shall be able to retain accurate records of training data, derived datasets, various versions of models, and downstream recipients. Further, an adversarial fine-tuning system can also be used, as developed by Deeb & Roger (2025). Their research on adversarial fine-tuning found that such a method could restore 88% of pre-unlearning accuracy for the information obtained during pretraining. While their research pertains to language models rather than genomic AI, it underscores the necessity for an unlearning procedure not to automatically be regarded as conclusive evidence of data removal. The SPAR project similarly recognizes the robustness of genomic-model unlearning as an area needing further scrutiny. Thus, India should advocate for proportionate verification, documentation of remaining uncertainties, and a clear distribution of responsibilities among relevant institutions, developers, and data fiduciaries.

A further step, as suggested by Sherkow (2025), is to provide secure computational environments that facilitate genomic analysis without necessarily transferring unrestricted copies of the underlying data to analytical services. In conjunction with the privacy protections outlined by Annan et al. (2025), this bolsters the case for considering data minimization, access controls, secure processing, and traceable records of data usage, thereby complying with the legal requirements as ascertained in the DPDP Act, ICMR guidelines and SAHI 2026. These privacy-by-design models shall help in enhancing, rather than substituting, the erasure procedures (Ginart et al., 2019).

Suggested Framework for Evaluating Genomic-AI Erasure Requests

These recommendations can be put into action through a six-stage evaluation framework that is rooted in Individual Autonomy and can aid in balancing innovation and Privacy

Stage 1: Identify the request and relevant legal obligation.

Clarify what the individual is asking for within the applicable legal provision i.e., whether the individual is asking to minimize the collection of data, collect only relevant data required, revoke consent, or erase the data altogether. Based on the request, make the relevant consent arrangements, see the retention requirements for the collection of data, and inform the individual of any statutory exceptions. It is also important to make distinctions wherever and whenever necessary between the right to erasure, withdrawal of consent, and a request to eliminate the influence of data from a trained model.

Stage 2: Map the genomic-data processing chain

Determine the original dataset, duplicates, derived datasets, training records, model versions, fine-tuned models, and known downstream recipients. Apply the necessary technical compliance and utilize available provenance and processing records to ascertain where the data has been utilized and which systems might be impacted.

Stage 3: Evaluate residual privacy and processing risks.

Assess whether relevant personal data is still being processed and whether the model or related systems pose reasonably foreseeable risks of disclosure, re-identification, or inference. Such evaluation should consider the structure of the AI model, the sensitivity of the data involved, and its intended application, without presuming that every trained model retains identifiable data.

Stage 4: Choose an appropriate response.

In accordance with the legal obligation and specific technical circumstances, the actions may include removing source records and duplicates, limiting further processing, addressing downstream datasets, employing machine unlearning, or retraining a model. The archiving of a model or retraining should not be an automatic process; likewise, the deletion of source data should not be automatically assumed to resolve all concerns regarding ongoing model-level processing.

Stage 5: Confirm and document the outcome.

Record the actions taken, the evaluated systems, the tests conducted, and the limitations of the verification process. In instances where unlearning is applied, assess its effectiveness through suitable testing proportional to the sensitivity of the information and the consequences of continued processing. Document significant uncertainties instead of making unfounded assertions of complete erasure.

Stage 6: Communicate the outcome and offer appropriate review.

Clearly explain the response, the actions undertaken, and any significant limitations or reasons for not performing a specific action, in line with applicable legal requirements. When necessary, provide a suitable review or escalation process for disputed or technically intricate cases.

Collectively, these stages link legal assessment, processing-chain mapping, technical responses, and accountability. They establish a structured method without presuming that every request necessitates model destruction or that every model inherently retains identifiable information.

CONCLUSION

The right to Erasure/ The right to be forgotten has become especially important and a conversation point in privacy jurisprudence since the utilization of personal data/ sensitive data for the training and development of AI systems. The one category of sensitive data that needs special protection is Genomic information. Utilisation of Genomic data for AI training poses the additional question of whether information derived from a deleted record may still be accessible, inferable, or influential in a trained model, unlike traditional data processing, where deleting a record may only solve the immediate instance of storage. Since genomic data is sensitive, it has the ability to expose information not just about the individuals but also about biological relatives. The efficacy of erasure is especially important to assess in the context of genomic data. For the purpose of this paper, we assessed the efficacy of India's data privacy framework to see whether it effectively protects the genomic information and the Right to erasure/ right to be forgotten. The DPDP Act, 2023, creates a statutory framework for erasing data, with certain limitations and conditions. Additional ethical and policy advice for responsible biomedical data use and AI governance may be found in the Strategy for Artificial Intelligence in Healthcare for India and the ICMR Ethical Guidelines for Application of Artificial Intelligence in Biomedical Research and Healthcare. However, after model training, the interaction between these instruments must be examined more closely in the event of an erasure request. The difference between removing source data and evaluating the information generated from it, in particular, raises concerns regarding execution, confirmation, and accountability. These concerns must be viewed in conjunction with the legal status of regulations that effectively mitigate such concerns. It is also important to make a distinction between enforceable law and a guideline, which essentially do not create a legally enforceable statutory duty. As a consequence, the present study indicates that the primary challenge lies not in the provision or enforcement of the right to erasure, but in determining how that right can be effectively implemented within the framework of training artificial intelligence systems on sensitive information such as genomic information. Technical approaches such as retraining and machine unlearning offer solutions; however, they should not be regarded as universally applicable until their suitability and effectiveness are assessed.

The research recommends a six-phase framework for assessing genomic-AI erasure requests. This framework intends to connect the legal entitlement of the Right to Erasure to a systematic implementation process by identifying the pertinent legal requirements, mapping the data processing chain, assessing residual risks, choosing suitable measures, documenting and validating results, and communicating decisions. It does not presume that every request necessitates model retraining or that complete technical erasure is always achievable. Instead, it emphasizes a justified, risk-aware evaluation supported by explicit accountability and corroborating evidence.

Finally, responsible data governance, technical evaluation, and legal coordination are necessary for successful data erasure in genomic AI. A thoughtful strategy must go further than the mere removal of files and consider if the actions taken are sufficient to handle the relevant legal obligations and privacy threats.

REFERENCES

1. Annan, R., Noland, J., Perkins, K., Yuan, X., Roy, K., & Qingge, L. (2025). Genomic privacy and security in the era of artificial intelligence and quantum computing. *Discover Computing*, 28(1), Article 108. <https://doi.org/10.1007/s10791-025-09627-w>
2. Borkar, J., Jagielski, M., Lee, K., Mireshghallah, N., Smith, D. A., & Choquette-Choo, C. A. (2025). Privacy ripple effects from adding or removing personal information in language model training. In *Findings of the Association for Computational Linguistics: ACL 2025* (pp. 18703–18726). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2025.findings-acl.959>
3. Cavinato, T., Hofmeister, R. J., & Kutalik, Z. (2026). Evaluating anonymized genome re-identification using polygenic predictions and its implications for data privacy [Preprint]. *bioRxiv*. <https://doi.org/10.64898/2026.06.10.731306>
4. Chandran, R. (2024). Collection of DNA and the creation of DNA databases in India: A proportionality analysis on whether there is a violation of genetic privacy. *Asia Pacific Journal on Human Rights and the Law*, 25(3), 282–308. https://brill.com/view/journals/aphu/25/3/article-p282_003.xml
5. Chen, J., Wang, W. H., & Shi, X. (2021). Differential privacy protection against membership inference attack on machine learning for genomic data. *Pacific Symposium on Biocomputing*, 26, 26–37.
6. Chourasia, R., & Shah, N. (2023). Forget unlearning: Towards true data-deletion in machine learning. In *Proceedings of the 40th International Conference on Machine Learning* (Vol. 202, pp. 6028–6073). PMLR.
7. Clayton, E. W., Evans, B. J., Hazel, J. W., & Rothstein, M. A. (2019). The law of genetic privacy: Applications, implications, and limitations. *Journal of Law and the Biosciences*, 6(1), 1–36. <https://doi.org/10.1093/jlb/lbz007>
8. Dara, M., Dianatpour, M., Azarpira, N., & Tanideh, N. (2025). The transformative role of artificial intelligence in genomics: Opportunities and challenges. *Gene Reports*, 41, Article 102314. <https://doi.org/10.1016/j.genrep.2025.102314>
9. Davis-Turak, J., Courtney, S. M., Hazard, E. S., Glen, W. B., da Silveira, W., Wesselman, T., Harbin, L. P., Wolf, B. J., Chung, D., & Hardiman, G. (2017). Genomics pipelines and data integration: Challenges and opportunities in the research setting. *Expert Review of Molecular Diagnostics*, 17(3), 225–237. <https://doi.org/10.1080/14737159.2017.1282822>
10. Deeb, A., & Roger, F. (2025). Do unlearning methods remove information from language model weights? *arXiv*. <https://arxiv.org/abs/2410.08827>

11. Dhaarani, R., & Reddy, M. K. (2025). Progressing microbial genomics: Artificial intelligence and deep learning driven advances in genome analysis and therapeutics. *Intelligence-Based Medicine*, 11, Article 100251. <https://doi.org/10.1016/j.ibmed.2025.100251>
12. Digital Personal Data Protection Act, No. 22 of 2023 (India).
13. Digital Personal Data Protection Rules, 2025 (India).
14. EY. (n.d.). Decoding the Digital Personal Data Protection Act, 2023. https://www.ey.com/en_in/insights/cybersecurity/decoding-the-digital-personal-data-protection-act-2023
15. Ginart, A., Guan, M., Valiant, G., & Zou, J. Y. (2019). Making AI forget you: Data deletion in machine learning. *Advances in Neural Information Processing Systems*, 32.
16. Guy, J., & Boonthum-Denecke, C. (n.d.). Forgetting by design: Testing the effectiveness of machine unlearning in right to be forgotten data deletion. *The International FLAIRS Conference Proceedings*. https://www.researchgate.net/publication/404577615_Forgetting_by_Design_Testing_the_Effectiveness_of_Machine_Unlearning_in_Right_to_Be_Forgotten_Data_Deletion
17. HBS AI Content & Learning, Harvard Research HBS, & HBS AI Affiliate Research. (2025, February 13). The myth of machine unlearning: The complexities of AI data removal. Harvard Business School AI Institute. <https://aiinstitute.hbs.edu/the-myth-of-machine-unlearning-the-complexities-of-ai-data-removal/>
18. Hein, Z. M., Guruparan, D., Okunsai, B., Che Mohd Nassir, C. M. N., Ramli, M. D. C., & Kumar, S. (2025). AI and machine learning in biology: From genes to proteins. *Biology*, 14(10), Article 1453. <https://doi.org/10.3390/biology14101453>
19. Huang, Y., Boja, E., & Auvel, J. G. (2026). Genomic relational privacy in AI era: A scoping review [Conference abstract]. *Cancer Research*, 86(7_Suppl.), Article 1397. <https://doi.org/10.1158/1538-7445.AM2026-1397>
20. Indian Council of Medical Research. (2023). Ethical guidelines for application of artificial intelligence in biomedical research and healthcare.
21. Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*, 31(3), 685–695. <https://doi.org/10.1007/s12525-021-00475-2>
22. Ju, Z., Xue, Y., Rud, A., Savatt, J. M., Lerner-Ellis, J., Rehm, H. L., Joly, Y., & Uberoi, D. (2026). Examining gaps in institutional policies for clinical genomic data sharing: A cross-jurisdictional study. *The American Journal of Human Genetics*, 113(9), 1801–1811.
23. Kohan, L. (2026). Artificial intelligence and genomic data privacy: Balancing innovation with security. *Molecular Biology Research Communications*, 15(1), 1–2. <https://doi.org/10.22099/mbrc.2025.54335.2217>
24. Lorenzo, D., Esquerda, M., Bofarull, M., Cusi, V., Roig, H., Bertran, J., Carrera, J., Torralba, F., Cambra, F. J., Vila, M., Garriga, M., & Palau, F. (2023). The reuse of genetic information in research and informed consent. *European Journal of Human Genetics*, 31(12), 1393–1397. <https://doi.org/10.1038/s41431-023-01457-y>
25. Mahmud, M. I., & Banerjee, T. (2026). Artificial intelligence in genomics: A comprehensive survey of methods, resources, challenges, and prospects. *Briefings in Bioinformatics*, 27(3), Article bbag229. <https://doi.org/10.1093/bib/bbag229>
26. Ministry of Health and Family Welfare. (2026). Strategy for artificial intelligence in healthcare for India (SAHI). Government of India.
27. National Human Genome Research Institute. (2022, January 12). Artificial intelligence, machine learning and genomics [Fact sheet]. National Institutes of Health. <https://www.genome.gov/about-genomics/educational-resources/fact-sheets/artificial-intelligence-machine-learning-and-genomics>
28. O'Connor, O., & McVeigh, T. P. (2025). Increasing use of artificial intelligence in genomic medicine for cancer care—the promise and potential pitfalls. *BJC Reports*, 3, Article 20. <https://doi.org/10.1038/s44276-025-00135-4>
29. Qureshi, S., Shaik, T., Tao, X., Xie, H., Li, L., Yong, J., & Jia, X. (2025). Exploring incremental unlearning: Techniques, challenges, and future directions. *arXiv*. <https://doi.org/10.48550/arXiv.2502.16708>
30. Royal, C. D., Novembre, J., Fullerton, S. M., Goldstein, D. B., Long, J. C., Bamshad, M. J., & Clark, A. G. (2010). Inferring genetic ancestry: Opportunities, challenges, and implications. *The American Journal of Human Genetics*, 86(5), 661–673. <https://doi.org/10.1016/j.ajhg.2010.03.011>
31. Sherkow, J. S. (2025). A sociotechnical approach to genomic data privacy: A comparative analysis. *International Data Privacy Law*. Advance online publication. <https://doi.org/10.1093/idpl/ipaf018>
32. Simon, G. E., Shortreed, S. M., Coley, R. Y., Penfold, R. B., Rossom, R. C., Waitzfelder, B. E., Sanchez, K., & Lynch, F. L. (2019). Assessing and minimizing re-identification risk in research data derived from health care records. *eGEMs*, 7(1), Article 6. <https://doi.org/10.5334/egems.270>
33. Sutthivana, (2026, April 22). Why AI cannot forget genomic data. Stanford Law School. <https://law.stanford.edu/2026/04/22/why-ai-cannot-forget-genomic-data/>
34. Thomas, M., Mackes, N., Preuss-Dodhy, A., Wieland, T., & Bundschus, M. (2024). Assessing privacy vulnerabilities in genetic data sets: Scoping review. *JMIR Bioinformatics and Biotechnology*, 5, Article e54332.
35. Tyagi, T. (2026, January 12). Rethinking the right to be forgotten in India's AI governance. Observer Research Foundation. <https://www.orfonline.org/expert-speak/rethinking-the-right-to-be-forgotten-in-india-s-ai-governance>
36. Venkatesaramani, R., Malin, B. A., & Vorobeychik, Y. (2021). Re-identification of individuals in genomic datasets using public face images. *Science Advances*, 7(47), Article eabg3296. <https://doi.org/10.1126/sciadv.abg3296>
37. World Health Organization. (2024). Guidance for human genome data collection, access, use and sharing. WHO.
38. Xu, Y. (2024). Machine unlearning for traditional models and large language models: A short survey. *arXiv*. <https://arxiv.org/abs/2404.01206>