PDF Download
3779299.pdf
21 January 2026
Total Citations: 0
Total Downloads: 199

Published: 08 January 2026
Online AM: 05 December 2025
Accepted: 25 November 2025
Revised: 23 July 2025
Received: 12 December 2024

Citation in BibTeX format

TUTORIAL

The Many Faces of Data Deletion: On the Significance and Implications of Deleting Data

IGNACIO MARCO-PÉREZ, University of La Rioja, Logrono, La Rioja, Spain

BEATRIZ PÉREZ, University of La Rioja, Logrono, La Rioja, Spain

ÁNGEL LUIS RUBIO GARCIA, University of La Rioja, Logrono, La Rioja, Spain

MARÍA ANTONIA ZAPATA, University of Zaragoza, Zaragoza, Zaragoza, Spain

Open Access Support provided by:

University of La Rioja

University of Zaragoza

ACM Computing Surveys, Volume 58, Issue 7 (May 2026)
https://doi.org/10.1145/3779299
EISSN: 1557-7341

The Many Faces of Data Deletion: On the Significance and Implications of Deleting Data

IGNACIO MARCO-PÉREZ, University of La Rioja, Logroño, Spain
BEATRIZ PÉREZ, University of La Rioja, Logroño, Spain
ANGEL LUIS RUBIO GARCIA, University of La Rioja, Logroño, Spain
MARÍA A. ZAPATA, University of Zaragoza, Zaragoza, Spain

Today, our data is not only stored on personal computers, but is managed by many devices, from cell phones or watches to smart TVs, and stored in remote repositories (usually referred to as "the cloud"). In this new context, defining what exactly "data deletion" is becomes a challenge, especially considering the many different scenarios in which it is becoming more increasingly important. This is the case, for example, of the "right to be forgotten" established by regulations such as the European General Data Protection Regulation (GDPR) or the deletion of data used as a source to feed machine learning processes, the long-term effects of which are very difficult to estimate. This work reviews the various terminology used when dealing with data deletion and analyzes the different fields and technologies to which it is related. We conclude by offering a structured discussion of key takeaways, lessons learned, and future research directions.

CCS Concepts: • **Information systems → Data cleaning**; **Information lifecycle management**; • **Applied computing**;

Additional Key Words and Phrases: Data deletion, unrecoverable data, data governance

ACM Reference Format:
Ignacio Marco-Pérez, Beatriz Pérez, Angel Luis Rubio Garcia, and María A. Zapata. 2026. The Many Faces of Data Deletion: On the Significance and Implications of Deleting Data. *ACM Comput. Surv.* 58, 7, Article 174 (January 2026), 31 pages. https://doi.org/10.1145/3779299

1 Introduction

Any system, any computer application, has to take a stance on the concept of *deletion*. By far, the most common is that deletion is an allowed option. For example, any operating system includes an option to delete files and/or folders. In operating systems with textual interface (via console), there is always a command (such as `DEL` in MS-DOS, or `rm` in UNIX-like systems) with different deletion options. In the graphical interfaces of this type of systems, the *garbage can* metaphor is normally used under different names (*Trash* in MacOS, *Recycle Bin* in Windows, etc.). A first notion

Ignacio Marco-Pérez, Beatriz Pérez, Angel Luis Rubio Garcia, and María A. Zapata contributed equally to this research.

This work was supported in part by Grant PID2024-155834NB-I00 by MICIU/AEI/10.13039/501100011033 and by ERDF/EU, and by project AFIANZA 2024/01 granted by the Autonomous Community of La Rioja.

Authors' Contact Information: Ignacio Marco-Pérez, University of La Rioja, Logroño, Spain; e-mail: ignacio.marco@alum.unirioja.es; Beatriz Pérez, University of La Rioja, Logroño, Spain; e-mail: beatriz.perez@unirioja.es; Angel Luis Rubio Garcia (corresponding author), University of La Rioja; Logroño, Spain; email: arubio@unirioja.es; María A. Zapata, University of Zaragoza, Zaragoza, Aragon, Spain; e-mail: mazapata@unizar.es.

ACM 0360-0300/2026/01-ART174
https://doi.org/10.1145/3779299

of "two-step deletion" appears in this metaphor: the first step is to send some content to the garbage can; the second, more "definitive" (see Section 3), is to empty the garbage can.[1] From there, any other file management protocol or application (FTP, WebDAV, Google Drive, Dropbox, to name just a few) also has a more or less sophisticated (or more or less directly linked to the underlying operating system) deletion method.

Thinking in terms of technical computer languages, most of them have some notion of deletion. Examples abound: the `remove()` library function in the C programming language (used to delete a file), the `remove()` methods used for example in the Java programming language (to remove items from `List` and `ArrayList` objects), the `Delete From` statement in the SQL language (used to delete rows in a relational table and "replicated" in the data languages of other data management systems), and so on. The more than well-known acronym `CRUD` that compiles the four basic operations in any data management procedure has *deletion* as its last constitutive element.

Raising the level of abstraction, it seems clear that any computer application has procedures allowing different types of deletion, from the most basic –through the *Backspace* and *Delete* keys of any keyboard–, to the most sophisticated embedded in complex programs such as spreadsheets, going through operations that all of us (as computer users) assume to be natural, such as the *Cut* option (together with its "sisters" *Copy* and *Paste*) or the *Undo* option. We can affirm that the availability of deletion is a characteristic element of computing itself, and that a significant part of the software engineering discipline is built around deletion management.

Nowadays, *data deletion* has become much more complex. We are even facing new computing paradigms in which, contrary to what we stated at the beginning of this introduction, erasing is not an option. This is the case of *blockchain* technology (more information about blockchain is given in Section 5), one of whose fundamental principles is based on the fact that the data it contains cannot be altered (in particular, it cannot be deleted), so the construction can be used as an unforgeable ledger [118]. Another technological example where deletion of information is not considered is the *Occurrence-centric approach* [28, 29]. In this proposal, the concept of *Occurrence Base (OcBase*) is defined, in which "the removal of any occurrence is prohibited. Even if an occurrence was registered by mistake, the occurrence must be stored in the base. In this case, the occurrence is marked by the user as "not valid" and the corresponding incidence is registered." This approach bears some resemblance to the notion of *valid time* included in *temporal databases* [142] (see Section 5).

In any case, the fact is that today data and files are distributed in a multitude of devices of a very diverse nature: cars and other motor vehicles, smart TVs and watches, mobile phones, tablets, laptops, personal computers, removable hard drives, IoT devices, remote data servers, and so on. The responsibilities regarding the information stored (and therefore, eventually erasable) in all these devices are different (and sometimes they are not even unambiguously assigned). To give an (apparently) simple example, the creation of a user account on any online platform can lead to the spread of data on a multitude of servers and devices. The fact that the account works correctly is proof of its successful creation, but, how can a user be sure that his/her data has been completely deleted if he or she requests the cancellation of the account? Who is responsible for guaranteeing the secure and definitive deletion of this information? And above all, who guarantees that such information can not be used in the future for unauthorized purposes? This kind of questions have led to the appearance in recent years of different regulations that, to a greater or lesser extent, try to provide users with more rights over the information they generate, being the European General Data Protection Regulation (GDPR) [33] and its "right to be forgotten" as the most relevant example of this type of legislation. In this new context, *data deletion* can no longer be considered only as an exclusively technical-computing operation, but something with ethical, social, and legal repercussions.

[1] This "two-step" notion is common to other user interface operations. For instance, they are named as *remove* and *delete* in https://adnanpuzic.medium.com/the-difference-between-remove-and-delete-7f6c1771e40f. (Last visited on December, 2025).

The objectives of this article are twofold. First, we intend to compile the multiple implications and repercussions of the notion of *data deletion* in a single source, through a three-dimensional analysis. The first dimension analyzed is the terminological one, reviewing the different terms used in the literature concerning deletion. As a second dimension, different fields of application and their particularities with respect to deletion operations are analyzed. The third and last dimension is the technological one, reviewing how different specific technologies approach data deletion. The second objective of this article is to help both researchers and practitioners to recognize that, in any data management task, deletion must be considered an essential function. To this end, we emphasize the importance of considering the three dimensions discussed (terminological, field of application, and technological), and provide a structured reflection —through a dedicated discussion section— summarizing key insights from our analysis, highlighting practical takeaways, and outlining future research directions to foster deeper awareness around data deletion.

The article is structured as follows. Section 2 presents related work. In Section 3 we review the complex terminology used to refer to the concept of *deletion*. Then we analyze in Section 4 how this concept is approached from different fields such as legal, ethical, scientific, and so on. The particularities of deletion with respect to different specific technologies are described Section 5. Section 6 discusses key insights and outlines future research directions. Finally, we present some conclusions in Section 7.

2 Related Work

Data deletion has attracted significant attention over the last decades, leading to a growing number of works dedicated to the subject. This trend underscores its rising relevance and complexity within contemporary information systems. This body of literature reflects a broad recognition that deletion is not merely a technical operation, but a multifaceted concern that spans multiple layers of system design, policy, and practice.

In this context, it is worth noting that much of this literature remains focused on isolated aspects, addressing specific concerns without engaging with the broader landscape of data deletion. For example, there are numerous resources (primarily web-based and often of a divulgative nature), that provide overviews of data deletion, often clarifying terminology and highlighting distinctions between commonly used concepts [14, 17, 61, 110, 145]. Among these resources, the *BitRaser Knowledge Series* [14] stands out, which offers a structured, though non-academic, treatment of data deletion, covering mainly terminology, standards, and compliance guidelines aimed at guiding practitioners and organizations. While these resources are useful for introductory understanding, they generally do not go beyond this conceptual level to examine specific application domains or analyze technical complexities involved in data deletion. On the other hand, there are specific works that, although providing valuable reviews on data deletion, tend to focus on specific technical environments or storage contexts rather than offering a broader, multi-domain perspective. For example, Leom et al. [75] explore remote wiping and secure deletion on mobile devices, with a particular emphasis on flash storage (or NAND flash memory). Their work is valuable for its forensic perspective and real-device experimentation; however, their study remains narrowly scoped to the domain of mobile device security. Similarly, Diesburg and Wang [27] conduct a comprehensive survey on *confidential* data storage and deletion techniques on non-distributed, single-user computing environments (such as laptops, thumb drives, or external hard drives). Their study focuses on ensuring irrecoverability of deleted data, limiting itself to providing critical insight into performance and usability trade-offs of secure deletion methods, especially in file systems and local media. In addition to these, Ahmad and Afzal [2] conduct a focused review on assured data deletion in cloud computing environments. Their work centers on the key requirements, existing approaches, and inherent limitations associated with achieving complete and irreversible deletion in cloud infrastructures, particularly under scenarios where users lose direct control over the physical location and replication of their data. While the article offers a useful overview of existing approaches (including secure overwriting, disk scrubbing, and cryptographic techniques), it is narrowly confined to cloud storage scenarios.

Other more recent contributions take a broader perspective on data deletion. Among them, the work of Tebernum and Howar [148] merits particular attention here, also because their objective aligns in part with the goals of our own study. In particular, they mainly investigate to what extent data deletion is reflected in existing research, and present a structured taxonomy of data deletion, organizing it along six key dimensions: what (what data is under consideration), why (why certain data should or must be deleted), who (influence of or impact on human actors), when (time-based or event-based deletion), where (physical or logical location), and how (methods, strategies, and level of destruction). While both Tebernum and Howar's work and ours aim at elevating the visibility and structure of data deletion within the data management discourse, they differ significantly in focus and scope. Their objective is to provide a foundational framework for understanding and professionalizing deletion as part of the data life cycle. In contrast, our work adopts a broader analytical perspective, aiming at consolidating the diverse implications and manifestations of data deletion across three dimensions: terminological, field-specific, and technological. Our contribution lies in emphasizing deletion as a cross-cutting and essential function in data management, and in drawing attention to the diversity of meanings, contexts, and constraints that shape it across different disciplines and technologies. On the other hand, Ramokapane and Rashid [123] propose the paradigm of *Explainable Deletion* (ExD), which focuses on improving transparency and accountability in data deletion processes and on facilitating users' understanding of such processes. To support this proposal, they conduct a state-of-the-art review centered exclusively on aspects that motivate the need for explainability —namely, the lack of clarity around nominal data deletion, the limitations of assured deletion mechanisms, and users' perceptions and practices regarding how data is (or is not) deleted— rather than providing a general overview of the existing body of literature concerning data deletion.

3 Terminology

There are many English words commonly used to refer to the action and effect of 'deleting' such as *remove*, *erase*, *clear*, *wipe out*, *eliminate*, *eradicate*, *obliterate*, *rub out*, *expunge*, *destroy*, or (even) *vacuum*, among others. The meaning of these words differs in small nuances (such as between *remove* and *delete*, which are often considered synonyms[2]), which causes them to be often used almost interchangeably in natural language.

When we limit the scope to Computer Science, and the purpose is to make a scientific use of the terms related to *data deletion*, a greater precision would be expected. However, there is no standardized or commonly accepted terminology. In the literature we can find similar or even the same terms referring to different concepts (by *different* we mean here from "slightly" to "completely" different), or the use of distinct terms referring to the same concept. Therefore, our intention is to make a broad review of the disparate uses given to these terms and concepts. Some of the terminological diversity that will be examined in this section is presented in Figure 1, which shows the differences at a first level (supplemented with more details in Figure 2).

3.1 Recoverable vs. Unrecoverable

A first key distinction concerns the recoverability of deleted data. Bakke et al. state that *data deletion* "refers to the assurance that deleted information is actually non-recoverable." [7] Garg et al. [44]

[2] https://wikidiff.com/remove/delete. (Last visited on December, 2025).

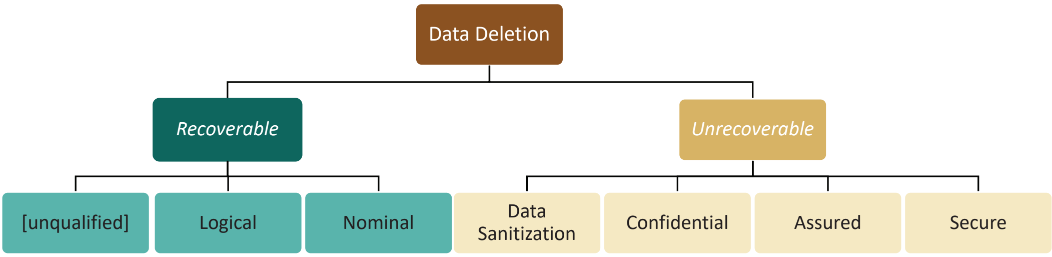

Fig. 1. Some of the diverse terms used regarding data deletion.

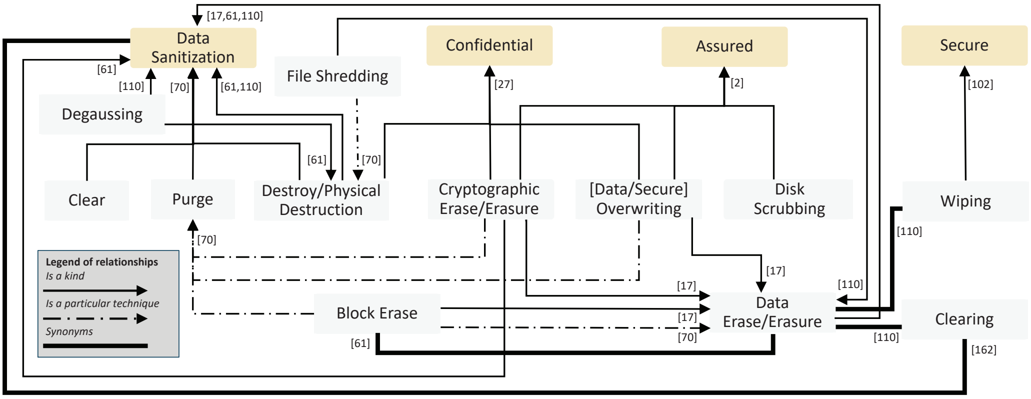

Fig. 2. Details of the terms used regarding *unrecoverable* data deletion.

propose a formalization of the notion of *data deletion* in the context of the 'right to be forgotten' (see Section 4), context in which it seems natural to assume that the *deleted data* should not be recoverable.

On the contrary, we can find sources in which the term *deletion* is linked to data that can be recovered. For instance, several informal sources, such as [17, 110, 145] use *data deletion* in this sense. In all of them, *deletion* (mainly concerning files within an operating system) is described as an operation that marks the space as free, but the data still exist on the hard drive, being easily recoverable by software. Another reference that shares this usage, although adding the adjective *logical* is [6]. For its authors, "Logical data deletion [...] does not guarantee purging of the data under deletion within a definite time frame. Rather, the data is marked as invalid; essentially, not accessible to external users. In practice, logically deleted entries are kept for arbitrarily long in the system, since the time to definitively delete the data (termed persistent deletion) depends on the state of the system." Taking a similar approach, Ramokapane and Rashid [123], use the term *nominal data deletion* to refer to the situations in which "the deleted item may no longer be visible, [but] it is technically not fully deleted and can still be recovered." Note how the (natural) idea emerges that *logical/nominal deletion* requires a lower operational level than *persistent/permanent deletion* (see also [70], that refers to a "given level of effort" in order to make data access infeasible). It is important to be very aware of this difference, otherwise problems arise such as those highlighted by Raquibuzzaman et al. [125]: "Although the state-of-the-art data deletion methods make the data inaccessible through standard memory interfaces, recent research efforts demonstrate that data is partially or fully recoverable."

There are other terms that do seem unequivocally linked to the idea that data are not recoverable. Figure 2 presents a selection of these terms, along with the relationships among them —such as "is a kind", "is a particular technique", or "synonyms"—as they appear in the literature (the figure also includes, in brackets, references from which these relationships are inferred). One of these relevant terms is *data sanitization*, defined as "the process of deliberately, permanently and irreversibly removing or destroying the data stored on a memory device to make it unrecoverable."[61] The term *data sanitization* is also used in a very relevant reference, that of the definition of DoD 5220.22-M [25], although with a slightly different meaning, that includes an idea of 'data protection': "Sanitization is the process of removing the data from media before reusing the media in an environment that does not provide an acceptable level of protection for the data that was in the media before sanitizing" (note the use of the word *removing* in this definition: in [163], and referring to [25], *data clearing* is defined almost identically to *data sanitization*, only replacing the word *removing* by *eradicating*). Another standard that refers to *(media) sanitization* is the publication SP800-88 Rev.1 [70], in which it is defined as "a process that renders access to target data on the media infeasible for a given level of effort", and mention should be made of the IEEE 2883-2022 standard for *Sanitizing Storage* [22]. In addition, there are several references ([14, 86, 104, 110, 138, 147, 163]) that provide definitions of *data sanitization*, either by referring to the above sources and/or using other expressions such as *data destruction*, *reliable remove*, *wiping* or *erasing*.

The other terms usually tied to unrecoverability are *secure data deletion* and *assured data deletion*. In these cases, the need to add an additional qualifier would seem to be an indication that for these authors, *data deletion* -unqualified- is linked to recoverable deletion. For instance, Reardon et al. [129] state that "secure data deletion is the task of deleting data from a physical medium [...] so that the data is irrecoverable. This irrecoverability distinguishes secure deletion from regular file deletion, which deletes unneeded data only to reclaim resources." In this regard, the book "Secure Data Deletion" by Joel Reardon [128] offers a detailed analysis of the topic, including descriptions of various mathematical techniques (such as B-trees and graph theory) that illustrate both the relevance and the complexity of the problem. In a similar line, Zheng et al. [175] states that "assured data deletion means that the outsourced data are permanently inaccessible to anybody upon requests of data deletion", and Ahmad and Afza [2] claim that "in some condition user wants to delete permanently his data. That should be unrecoverable in very reliable manner after a particular time. Assured deletion is very important." The disparity of terms used is expressed by Leom et al. [75], who say that "Secure deletion is sometimes referred to as forgotten, erased, deleted, completely removed, reliably removed, purged, self-destructed, sanitized, revoked, assuredly deleted, and destroyed in the literature." Even so, this list is not exhaustive. For instance, Gnatyuk et al. [48] uses the term *guaranteed data deletion* to refer to the same concept, further illustrating the terminological diversity surrounding this notion.

The term *data erasure* is also often related to unrecoverable deletion. It is defined in [145] as "the process of overwriting data so that it can no longer be recovered." In turn, in [110] it is said that "*data erasure* is sometimes referred to as *data clearing* or *data wiping*", and specifies it as a particular case of *data sanitization*. This classification is shared by [17], and also by [61]. Another different term is *block erase*, that is used as a synonym of *data erasure* in [61], as a particular case of *data erasure* in [17], and as a specific technique of *purge* by Kissel et al. [70] (see next subsection for details).

3.2 Specific Cases of Unrecoverable

Several sources present *classifications* when referring to non-recoverable deletion. These sources have in common that they specify three different types and usually include a notion of *physical destruction*. However, the classifications are different in some aspects, both in the terms used and in the definitions presented.

As a first example, Kissel et al. [70] divide *data sanitization* into three categories: *clear* ("based on overwriting [...] using standard Read/Writing commands"), *purge* ("employs techniques such as Overwriting, Block Erase and Cryptographic Erase that use specific commands") and *destroy* (with "several techniques, such as shredding, disintegration, melting, incineration, and so on., to destroy the storage media physically"). However, in [110] it is said that "data sanitization is achieved through three major methods: physical destruction, degaussing, and data erasure", while in [61] it is stated that "there are three methods to achieve data sanitization: physical destruction, cryptographic erasure, and data erasure".

Ahmad and Afza [2] put forward three proposals for *assured data deletion*: *secure overwriting*, *disk scrubbing* and *cryptography*. This last division is similar to the one proposed by Diesburg and Wang [27] for their term *Confidential data deletion*, with the variation that they use the terms *data overwriting*, *physical destruction* and *encryption with key erasure*, respectively. Finally, [17] classifies three types of *data erasure*: *overwriting*, *block erase* and *cryptographic erase*.

Apart from these classifications, we also focus on a few other terms related to unrecoverability. One of these terms (and one that has appeared previously) is *data wiping*. In [110] it is stated that it is "one of erasure methods. It is often used to erase data on a large scale." In [17] it is described as an example of partial data sanitization (although the author does not define this term). Ölvecký and Gabriska [104] point out that there are several data wiping techniques, such as the aforementioned DoD 5220.22 M, or NCSC-TG-025, among others. Oh et al. [102] indicate, focusing on the type of deleted information, that "the purpose of data wiping is secure deletion of undesired and classified files."

Another term is *File shredding* which is cited in [14], referring to [70], as a specific technique for *physically destroying* the storage media. In turn, in [17] it is included (without giving an explicit definition) as an example of *partial data sanitization*. In [110] *File shredding* is identified as a way of *data erase*, at the same level of *data wiping*. Another technique already mentioned is *degaussing*, which consists of exposing the media to a magnetic field, and is therefore included in [110] as one of the three methods of *data sanitization*, and in [61] as a particular case of *Physical destruction*. The terms *formatting*, *reformatting* or *factory reset* (this last one usually linked to mobile devices) are related to different ways of physical deletion. For instance, they are considered in [17] as examples of *partial data sanitization*.

To complete the overview of the different cases linked to unrecoverability, we give some details about *cryptographic erasure*. According to [61] it is a term "used interchangeably" with *crypto erase* (thus it is used in [110]), and it is a method of *data sanitization* consisting of "the process of using encryption software [...] on the entire data storage device, and erasing the key used to decrypt the data." In turn, in [17] it is stated that *cryptographic erase* is a kind of *data erasure*, while Diesburg and Wang [27] call it *Encryption with Key Erasure* in their classification of ways of *Confidential data deletion*. Finally, Politou et al. [119] mention *cryptographic erasure* in the context of the 'right to be forgotten', although they specify that "this method actually deactivates the personal data in question, rather than deleting it" (which would seem to go against the idea put forward by other authors that *cryptographic erasure* does not allow data recovery). Due to its inherently complex and mathematically rigorous nature, a detailed analysis of cryptographic erasure falls outside the scope of this work, which cannot accommodate the depth of detail such a topic requires. Nonetheless, it is worth briefly mentioning different solutions proposed in this field, such as *ErasuCrypto* [76] (which integrates 'erasure' and 'cryptographic' methods, applying them to solid state drives, SSDs), *SADUS* [169] (which tests its effectiveness through coercion attacks on Android devices), *STM Shredder* [48] (which includes a pseudo-random sequence generator), and *KDE* [164] (which is a new encryption algorithm used to build a secure deletion scheme referred to as SDDK).

3.3 Other Terms

There are a number of other terms that, without referring exactly to the concept of *deletion*, have some relation to it. The first of these terms is *data hygiene*, described in [61] as "the process of ensuring all incorrect, duplicate or unused data is properly classified and migrated into the appropriate life cycle stage for storage, archival or destruction". *Data hygiene* is also referred to as *data cleaning* [59], *data cleansing* [57] or even *data scrubbing* [20], and it is recognized as a key element of the machine learning success.

A very specific term is that of *vacuuming*. In the context of temporal databases, it is referred to as "physical removal of data" by Skyt et al. [140]. As a curiosity, that work points out that "For emphasis, we will use "delete" for logical deletion and "remove" for physical removal throughout the article".

Another important term is *deduplication*, that refers to "the automatic elimination of duplicate data in a storage system." [113] Deletion of duplicate data, in addition to the obvious advantages in terms of recovering available space, includes a number of challenges of its own, depending on the different storage media used.

Finally, a very relevant term regarding the management of personal data is *anonymization* [95]. One of the more used techniques for *data anonymization* is *data masking*, that following Winkler [163] "is intended to remove all identifiable and distinguishing characteristics from data in order to render it anonymous and yet still be operable." This author also claims that other synonyms for *data masking* are *data obfuscation*, *de-identification* and *depersonalization*. However, Kalaiselvi and Yoga [64] state that "Data obfuscation (DO) is a form of data masking [...] DO is also known as data scrambling and privacy preservation". In any case, what is clear is that everything related to personal data protection and control [139] has important repercussions at the social level (ethical, medical, environmental, etc.), which makes that scientific publications in fields such as law also have to take into account seemingly technical terms as *data anonymization* or *data masking* [157].

4 Data Deletion Fields

Having introduced terms and definitions related to deletion, we will now analyze the various issues and implications that arise in the different fields involved. When perspectives on different areas are included, even contradictory aspects appear, such as the need to keep the data for the reproducibility of the experiments and the 'right to be forgotten'. As a consequence, there is no single, definitive solution, but rather a compromise between different alternatives must be sought. As graphical support for this section, Figure 3 provides an overview of the different fields involved in data deletion, alongside the relevant technologies, which will be addressed in detail in Section 5.

4.1 Legal

Legal regulations of different countries establish, under certain circumstances, the right of individuals to request personal data deletion for preserving their privacy. One of the most classic rights is linked to an individual's judicial or criminal past, and establishes *the right to oblivion of the judicial past*. "It is justified by faith in a human being's capacity to change and improve as well as on the conviction that a person should not be reduced to their past." [24] As explained in [24], when this right conflicts with the right to information (also related to freedom of the press), time is often used as a key criterion to resolve the tension. In these cases, the right to information usually prevails when judicial decisions are considered newsworthy. However, as time passes and journalistic interest fades, the right to oblivion tends to prevail. In these cases, usually, the judicial case can be mentioned, but without naming individuals. In any case, regardless of the time that has elapsed, historical relevance and public interest can be considered so that the right to information overwrites that of oblivion of the judicial past.

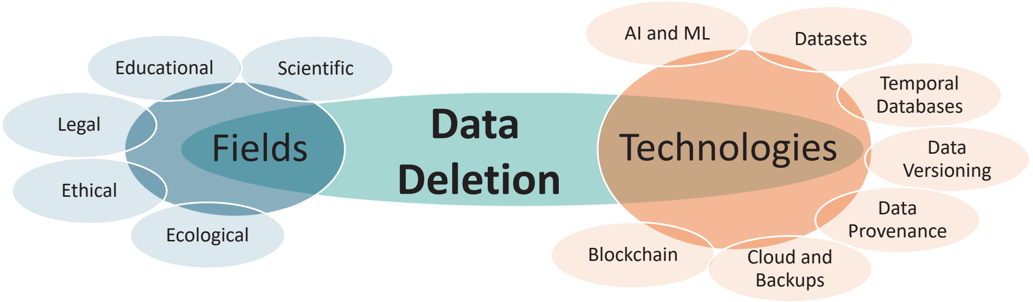

Fig. 3. Some fields and technologies concerned by data deletion.

Other regulations, such as *the right to be forgotten* (in EU's GDPR [33]), *right to delete* (in California's CCPA [60] and CPRA [149]), or *deletion right* (in Virginia's VCDPA [156]), are not restricted to judicial information but cover any personal data. These laws regulate persistent deletion of user data on-demand and in a timely manner [6]. Two key aspects of these laws, that pose significant technological challenges are their retroactive nature and the requirement that personal data be erased "from every data controller" who is processing the data and not only from the one who processed the data in the first place [117]. As a consequence, companies are forced to provide processes and mechanisms to be able to fulfill users' data deletion requests promptly and effectively [123]. The goal is to provide individuals with control over their personal data counteracting the public and easy availability of data through the Internet. It must be noted that the right to be forgotten (as well as the right to oblivion of the judicial past) conflicts with other laws, so it is necessary to carry out a case-by-case analysis. On the one hand, a balance must be struck between privacy and freedom of expression [73, 158]. On the other hand, reliability, integrity, and transparency principles must be followed by institutions (for instance, financial) so that data records may have to be kept in non-volatile storage [117].

Another type of oblivion right, related to Internet search engines, is *the right to delisting* (also called *right to delinking* or the *right to de-indexing*) [158] according to which "people, under certain conditions, have the right to have search results for their name delisted" [73], removing the links and not the information itself [117]. As noted by [117], the information is still retained, but access to it is more difficult since users have to know where to look for it. As a consequence, since the disputed information is not censored, "the decision was ultimately not about the fundamental balance between privacy rights and expression rights when dealing with personal information over the web." For example, since 2014, Google has received more than 7 million URLs delist requests under European privacy law and has delisted around 52%.[3]

It is also worth mentioning the identification-related regulation, such as the European Digital Identity Framework [35], also known as eIDAS2 (for being a modification of Electronic Identification, Authentication and Trust Services Regulation -eIDAS). This framework aims to "provide people with control over their online identity and data", in line with the Digital Decade Policy Programme 2030 [34], that sets the objective of developing "a trusted, voluntary, user-controlled digital identity that is recognised throughout the Union and allows every user to control their data in online interactions". Related to data deletion, eIDAS2 sets out that users should be empowered to "*request the erasure* of their personal data in a user-friendly and convenient way, under the sole control of the user, while enabling *selective disclosure* of personal data." In particular, European Digital Identity

[3] https://transparencyreport.google.com/eu-privacy/, for (Last visited on December, 2025).

Wallets (electronic identification means for secure management of person identification data) shall enable the user to easily request the erasure of personal data pursuant to the right to be forgotten of GDPR. Furthermore, selective disclosure empowers the user to disclose only such personal data as is necessary for the provision of the requested service, reinforcing data minimization principle. As a consequence, to the extend that the amount of data is reduced, deletion becomes easier.

Legislators deliberately specify the deletion requirements in a technology-agnostic approach, allowing their adjustment to future technical innovations [117]. As a consequence, the implementation of these laws, in the digital environment, is not a straightforward task, becoming burdensome or even impossible in some scenarios [117]. For example, "the need for data to be deleted from all possible sources in which they reside [...]. It is evident that the enforcement of this right would pose major technical issues due to the practicalities involved in knowing all the controllers who are processing the personal data in question." [117] Other implementation challenge described by Mangini et al. [83] is the removing of personal data from archives or backups since it is an arduous and difficult task. Cryptographic erasure is considered a suitable solution, but the problem is that "we are not deleting data but rendering them unusable" and "an attack on the key storage server will allow a malicious user to retrieve the keys and decrypt the data." [83]

Furthermore, organizations can be fined for non–compliance with the data deletion regulations. For this reason, deletion–compliant data systems, using different technological resources, have been proposed. For example, an extended query language for timely persistent data deletion is presented by Athanassoulis et al. [6], and a blockchain-based scheme for verifiable data deletion is provided by Yang et al. [168]. In addition, one strategy that makes it easier for organizations to comply with data deletion regulations is to minimize the amount of data collected. For example, *selective disclosure* property of Verifiable Credentials (VCs) [122] enables users to reveal only partial information extracted or derived from their credentials [63], thereby reducing the amount of data shared.

Finally, as another legal issue related to the *right to be forgotten*, we must mention the field of *digital data forensics*, which is often used as a result of a judicial or police intervention, and which on many occasions pursues the recovery of deleted data. It is a field with a long history in the relationship between law and computing, for which numerous challenges have been presented [66, 155], for instance in areas such as cloud [85], mobile [8] or IoT [68]. Regarding both smartphone and IoT technologies, it is worth mentioning the framework *DelSec*, which is presented precisely as an 'anti-forensics data deletion' solution. The number and types of challenges related to data forensics continue to grow as new technologies emerge: for example, Onik et al. [105] present a systematic review of digital forensics in instant messaging applications (WhatsApp, Telegram, Signal, among others), and Feng, with several authors in two separate works [37, 38], presents different artificial intelligence techniques for detecting the deletion of video frames in forensic contexts.

4.2 Ethical

Very relevant ethical debates have been raised about the management of digital data, giving rise to critical positions that, for example, claim that "machine learning models can reproduce or amplify unwanted societal biases reflected in training datasets." [45] The issue regarding deletion is that even after this biased data has been deleted, it remains "as residues in algorithmic models, or as data set derivatives." [150]

Data deletion is also addressed from an ethical point of view in relation to death. Artefacts that remain in social network services (SNS) after the death of users are considered as a significant material instantiation of persons in [146]. In this sense, deletion of a person's social media presence is seen as a second death and it is claimed a corresponding reason not to delete the SNS profiles of the deceased. Related to this, there is a burgeoning digital afterlife industry that raises difficult ethical concerns. As regards deletion, two questions are proposed by Öhman and Floridi [103]: (1) "If not deleting [the digital remains], what would make the cost of storing billions of dead profiles financially viable?" and (2) "if the choice will be to delete profiles, what would the selection process look like?"

A large-scale, multi-platform analysis of end-of-life support is performed in [31], concluding that "most platforms offer minimal functional support and that many end-of-life needs are impossible to meet or rely on unsanctioned workarounds." The platform policies for post-mortem data management analyzed in this study are varied, for example, Twitter (now X) allows bereaved to request the removal of the decedent's X account, however, on the opposite side, LINE (Japanese freeware social network) has no method to delete the account of a deceased, even if login information is known. As other examples, Facebook's Legacy Contact allows account holders to choose someone to manage their profile after they have died [42], and Google's Inactive Account Manager deletes inactive accounts after a set time [46].

These ethical aspects about digital remains of dead people has as a counterpart the right to be forgotten by protecting data privacy, mentioned in the legal field. Therefore, arguments from different fields of reference must be taken into account, showing that "deleting [information] generally requires careful consideration and a compelling reason, and the permissibility of some kinds of deletion are often complex or unclear." [55]

4.3 Ecological

Computing has to address the planet-scale limits and cannot assume that ever-increasing production and consumption will continue [98]. Stored data has been growing every year, and thus they consume increasing natural resources. Specifically, it has been estimated that worldwide enterprise data storage will continue to grow at a 27% compound annual growth rate [3], which will result in a great need for resources and energy. We will soon be facing limits to storage [55], so that we shift from the problem of what to save to the problem of what to erase, needing a better understanding of which data are worth preserving [40]. Thereby, although deletion has a broadly negative perception in comparison to preservation, there is a growing necessity of data deletion for the sake of environmental sustainability [55].

Environmental and sustainable issues about the previously mentioned digital remains of deceased users have also to be stated. According to Welsh [161], life is inherently ephemeral, and, with the goal of providing a sustainable model, digital traces might find a way to replicate that, allowing for and even encouraging decomposition and disappearance. In this way, this new perspective, in addition to those mentioned above, must be taken into account when making decisions about the maintenance or deletion of stored data.

Another aspect related to this field is the electronic waste (e-waste) generated by the physical destruction mentioned in Section 3. E-waste is growing every year, with a dangerous environmental impact if it is not treated safely. For example, incineration (mentioned previously as a type of physical destruction) "releases harmful chemicals in the atmosphere due to burning, and the residues left can be hazardous. It also contributes to global warming." [65]

Due to the negative environmental impact of data storage —such as carbon footprint and e-waste generation— the IT industry has implemented several strategies to mitigate these effects. For example, Amazon Web Services is committed to reducing carbon emissions "using clean, renewable energy sources to power its data centers." [127] Similarly, Google "prioritizes the reduction of e-waste through initiatives like product refurbishment and recycling programs, emphasizing a circular economy approach." [127] Despite these initiatives, data centers still consume significant energy to store large volumes of unstructured and inert content (commonly referred to as "dark data"), which will most likely never be used. In such cases, the most sustainable option would be to delete it [3].

4.4 Scientific

Accountability and transparency must be achieved in all scientific research results. One way is to be able to support method reproducibility providing enough detail about study procedures and data so that the same analysis and results could be repeated. That is, a researcher could duplicate the results using the same materials [49]. The problem is that, at the same time, the privacy of individuals must be preserved and their consent must be respected [32]. Therefore, "the difficulty is in finding a compromise between the conflicting needs of anonymity and of preserving data for audit or further analysis." [89]

For example, clinical trial datasets contain very detailed information on each participant. Risk to patient privacy can be mitigated by data reduction techniques. One recommendation is generating de-identified datasets removing or recording potential indirect identifiers [152]. However, de-identification has two problems. On the one hand, "an excessive application of such techniques may pose a public health risk if misleading results are produced." [152] On the other hand, there is a wide spectrum of human characteristics that enable re-identification, and powerful re-identification algorithms have been proposed [97].

In addition, although consent for scientific research studies has been given by users, they can later revoke their consent. In this situation, consent for past studies is still valid, so that, in order to reach a certain degree of method reproducibility, "minimum necessary meta-information must be kept in the provenance module for past studies that were conducted and based on now deleted information." [32]

As an alternative, "using synthetic data can help to avoid ethical and legal issues, in particular breaching the privacy of real users and the need for institutions to adhere to strict data protection regulations." [11]

4.5 Educational

Educational institutions store and manage personal data of their students. Taking into account that at some point the students will leave the institution, "there should be transparent policies about how long data can be held for and what the process is for handling requests for deletion of data." [141] However, as in other fields, there are different aspects that need to be analyzed. "The guarantee that personally identifiable data are deleted when a student leaves an institution may also have a positive impact on student trust, but at the same time, keeping student data will be helpful for the university to refine its analytics models, track the development of student performance over multiple years and cohorts or simply for internal or external quality assurance processes." [109]

For example, the data protection policy at Uppsala University[4] stated that students "have the right to have [their] personal data erased from Uppsala University's systems as long as they are not an official document". Specifically, "the right to erasure is severely limited by the regulations on official documents and by the requirements of research or study documentation."

5 Related Technologies

After examining the various fields in which data deletion plays a critical role, we now turn to the technological dimension (see Figure 3). This section analyzes how specific computer science technologies relate to the concept of *deletion*. In some cases, the relationship is of an intrinsic nature, in the sense that the technology itself has certain characteristics that link it directly to *deletion* (or to the *unfeasibility of deletion*): temporal databases, data provenance, or blockchain are examples of such technologies. In other cases, the relationship of technologies with *deletion* has

[4] https://www.uu.se/en/about-uu/data-protection-policy. (Last visited on December, 2025).

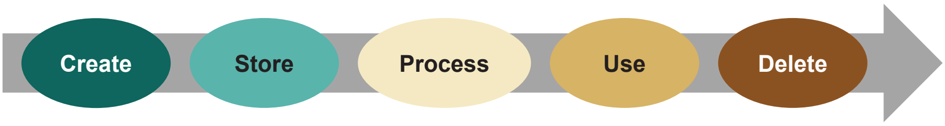

Fig. 4. Basic data lifecycle management.

more to do with the repercussions surrounding their use, with artificial intelligence and machine learning being the most prominent cases.

Given that data deletion encompasses various requirements and context-dependent approaches, the metrics used to evaluate the effectiveness of data deletion often vary according to the underlying technologies. Some approaches to this issue are framed within the aforementioned domain of data forensics [53], using as a metric the number of data points recovered by specific recovery algorithms. At the same time, a significant body of literature focuses on providing verifiability guarantees, as defined in each study, often designed based on cryptographic knowledge. For instance, Xu et al. [166] propose a verifiable scheme for secure data deletion in cloud storage, which incorporates an efficient verification algorithm as a core component of the model.

Before analyzing particular technologies, it is worth referring the concept of *data lifecycle* and its management, since it is a feature that is to some extent transversal to any specific technology or term (for example, data lifecycle management is described in the context of *data sanitization*, and also when defining *data hygiene*, in [61]). There are countless versions of the data lifecycle concept, with variations in both the name and the number of phases included in it. Similarly, there are numerous articles in the literature that have highlighted the importance of data lifecycle management analysis. Among many others, we can cite Ofner et al. [101], who work on the specific concept of *master* data lifecycle, Polyzotis et al. [120] with a survey on the challenges of lifecycle management in machine learning, or Rahul and Banyal [121], dealing with the lifecycle management in big data analytics. In Figure 4, we have reproduced a basic version of the lifecycle, including five phases, with *Deletion* being the last one. However, to consider deletion as "the last game" is described as one of the most frequent data lifecycle management mistakes, since "Already-disposed data might contain information that may be significant later",[5] which in a way brings us back to the previous discussion on recoverability or irrecoverability of deleted data.

5.1 Temporal Databases

One of the first technologies in which the concept of deletion took on a new meaning was in *temporal databases* [142]. Through the notions of *valid time* and *transaction time*, certain data that in a non-temporal database would simply be deleted (because it is considered that they should no longer be stored), are still retained in a temporal database (accompanied by the corresponding validity time stamps).

As is evident, this type of database could be ever-growing, and in addition to the (environmental) problem already mentioned in the previous section concerning the limits of storage, it was already recognized thirty years ago that indefinite retention of data could lead to both legal and management issues, which led to the use of the notion of *vacuuming* in temporal databases [140].

Much more recently, it has again become clear that the relationship of *time* to the storage (and thus deletion) of data is relevant. Thus, in [137], "Storing Data Forever" is enunciated as one of the "seven GDPR sins." Specifically, that reference notes that "all personal data must have a time-to-live (TTL)" and that "while conceptually clear, timely and guaranteed removal of data is challenging in practice."

[5] https://spanning.com/blog/4-data-lifecycle-management-mistakes-businesses-should-avoid/. (Last visited on December, 2025).

5.2 Data Versioning

Data evolve over time, and several reasons, such as to support the reproducibility of scientific experiments, make it necessary to track the different versions of data over time [13]. Whereas temporal databases restrict validity of each tuple to a time range, data versioning has a higher granularity, tracking the history of whole tables [67] and giving support for a branched network of versions [58, 134]. As a consequence, when data is deleted in a version, that information does not really disappear since it remains in previous versions.

Furthermore, since many versions can be similar to each other, with the goal of eliminating redundant storage, several versions are stored jointly enriching the data with versioning information. For example, Bhardwaj et al. [13] propose to extend each record with a deleted bit to track whether the record is active in a particular version. Similarly, the proposal of Schüle et al. [134] is that each branch maintains a bitmap for every table, denoting each tuple's visibility. Besides, a garbage collection to remove versions that are no longer contained in a branch is proposed in [134].

5.3 Data Provenance

Following the W3C perspective, *data provenance* is "information about entities, activities, and people involved in producing a piece of data or thing, which can be used to form assessments about its quality, reliability or trustworthiness." [92] The implications of deletion in relation to data provenance are clear, in at least two ways.

On the one hand, provenance information is metadata, and therefore data, which raises the issue that it is susceptible to deletion. Therefore, the need arises to establish policies that define under what circumstances the deletion of the provenance information must be carried out [15]. On the other hand, since the purpose of provenance is to trace the "existence of a piece of data", should provenance information be preserved even if such data is deleted? In [93] we can find a possible answer to this question, since its author states that "in some cases, provenance must persist even after the data it describes has been removed."

In the current contexts of personal data protection, one evidence of the relevance of provenance information management is the intense research that is recently being carried out on this topic [107, 108, 130, 153].

5.4 Blockchain

One of the key features (probably the most distinctive of all) that guarantees the integrity of blockchain technology is its immutability. Transaction data stored in a blockchain are not modifiable and cannot be deleted. However, this technology is still under development, so it is natural for alternative proposals to appear.

Thus, a very recent comprehensive review of the different challenges and solutions that exist in the literature around the concept of blockchain mutability is presented in [118]. The first concrete proposal around the idea of a blockchain that can be modified is the notion of *redactable blockchain* [5]. According to its authors, there are several main reasons for the need for this concept: first, there are blockchains such as Bitcoin that "contain child pornography, improper content, and material that infringes on intellectual rights" (while it may seem shocking to consider that Bitcoin could host content that is far away from its primary purpose as a decentralized payment system, this fact has been reliably proven through both quantitative and qualitative analysis [52, 87]); second, there may be performance or scale problems when using smart contracts and overlay applications; and third, there is a conflict between the use of blockchain to store personal data and the "right to be forgotten". Naturally, this last reason has been explored by other authors, such as Koscina et al. [10] in the context of the use of blockchain in the healthcare industry, or Zyskind et al. [178], who precisely propose the use of blockchain as a method of protecting personal data (in fact, Politou et al. [117] point out that the solution proposed in [178] is a “very good candidate for implementing the “right to be forgotten” requirement specified in the GDPR”).

There are other relationships between blockchain and the concept of deletion. For example, a deletion scheme for the cloud that considers the repercussions of such deletion on blockchain systems is presented by Khanboubi et al. [69]. In an almost opposite sense, Yang et al. [168] propose a method that uses blockchain to certify that deletion in the cloud has indeed been performed. More in general, Zou et al. [177] analyze the relationships between blockchain and cloud, as well as their associated challenges around data integrity and privacy protection (which as we have already discussed are intimately connected to deletion).

5.5 Cloud and Backups

The references to “the cloud” at the end of the previous section make it appropriate to expressly mention the impact of deletion with respect to this “technology”. Ahmad and Afza [2] state that once a cloud user decides to delete his/her data, it must be ensured that the data is actually deleted from all cloud storage sources, and that no copies remain in any part of the cloud storage infrastructure. The authors also suggest making a review of the mechanism of *assured data deletion* in this context. The importance of the cloud deletion process is recognized by companies such as Google itself. Note how several of the terms we have described in Section 3 are used in the official documentation on *Data deletion on Google Cloud* [50] and *Google’s statement on deletion and retention* [51]: “When you delete your Customer Data, Google’s deletion pipeline begins by confirming the deletion request and eliminating the data iteratively from application and storage layers, from both active and backup storage systems. [...] Logical deletion occurs in phases, beginning with marking the data for deletion in active storage systems [...] Successive compaction and mark-and-sweep deletion cycles [...] serve to overwrite the deleted data over time. [...] Cryptographic erasure is also used to render the deleted data unrecoverable.” We must also refer to the duration of the deletion process (in the particular case of the Google cloud): “Deletion from active systems typically completes within about two months of the deletion request. Finally, Customer Data is removed from Google’s long-term backup systems, which preserve snapshots of Google systems for up to six months.”

As can be deduced from the statements presented, the idea of cloud is strongly linked to the idea of backup (since in order to ensure data availability and quality of service in the cloud it is necessary to have different replicas of such data). When it comes to deletion, the existence of backups is an added difficulty, since for the deletion of some data to be definitive, it must be guaranteed that all existing copies of the data are deleted [83]. Note also that contrary to the possible preconceived notion that a backup is performed on large volumes of data, the fact that very specific data needs to be erased on backups makes it necessary to implement fine-grained operational techniques. This is reflected in the existence of studies comparing the behavior of different databases when deleting information from backup copies [74]. As expected, some authors have also observed that there is an uneasy relationship between backups and the “right to be forgotten” [119].

5.6 Artificial Intelligence and Machine Learning

There is also a substantial body of work addressing the relationship between Artificial Intelligence (AI) and the “right to be forgotten”, and consequently, the association between AI and data deletion, particularly concerning the application of Machine Learning (ML) techniques. In this context, as proposed by Ginart et al. [47], data deletion applied to ML is defined as the “goal that after a specified datapoint, x, is deleted, the resulting model is updated to be indistinguishable from a model that was trained from scratch on the dataset sans x.” Achieving this objective is often quite challenging. For example, Fosch-Villaronga et. al [41] state, referring to *data deletion*, that “this seemingly simple issue poses many practical problems in actual machine learning environments. In fact, “data deletion” requirements can be considered to actually border on the edge of impossibility.” In their analysis, Dam et al. [23] underscore how challenging it is to achieve this goal, stating that “the impact on ML classification performance highly depends on the number of deleted records, the specific characteristics of the dataset and which attribute values are most important for the classification.” Most of the works focus on finding technical and specific solutions to the problem that “for many standard ML [machine learning] models, the only way to completely remove an individual’s data is to retrain the entire model from scratch with the remaining data, which is often computationally impractical” [47]. For instance, Malle et al. [82] address the problem by applying machine learning on perturbed knowledge bases, Schelter [133] performs a decremental update on already existing ML models to forget a user’s data, and Izzo et al. [62] propose an approximate deletion method for linear and logistic models whose computational cost is also linear.

Everything surrounding data deletion in relation to ML has proven to be of enormous relevance, since the term *machine unlearning* (*MU*) has recently been coined to refer to mechanisms that allow the removal of data from a model without having to retrain it in its entirety. In fact, several comprehensive studies on this new sub-discipline are already publicly available [99, 136, 165, 172]. One example is the work by Rawat et al. [126], who delve into the application of Bayesian unlearning using estimation techniques. They tackle the task of determining new parameter values in neural networks following data deletion. Similarly, Cao et al. [16] introduce a method based on the projection residual method using Newton iteration. This approach is effectively employed in linear regression models and neural networks, providing an additional perspective on the implementation of MU.

While attention is typically focused on variations in model accuracy and recall following unlearning, the assessment of data deletion effectiveness —usually focused on contrasting the results with the retrained model— remains an underexplored issue [160]. The manner in which the trade-off between these two aspects is addressed is highly context-dependent, and various approaches have been proposed. For example, Wang et al. [160] propose a model difference simulation scheme based on influence function theory to generate the unlearning model difference. Vidal et al. [154] explore the use of *Explainable AI* (XAI) techniques to verify data removal, introducing novel metrics such as Heatmap Coverage and Attention Shift. Building on this, *Membership Inference Attacks* (MIAs) —which aim at determining whether a specific data point was used during the training of a target machine learning model [56]— have become a standard tool for evaluating MU techniques, often conceptualized as *backdoors*. A common approach involves introducing perturbations to input data in both the original and unlearned models and analyzing output discrepancies to detect deleted instances. This technique has shown that even MU models that output only labels (excluding prediction probabilities) remain vulnerable to MIAs [18, 79, 80]. Several studies investigate these emerging vulnerabilities and propose new defenses against such attacks, even for models explicitly designed to support data deletion [18, 30]. Some unlearning strategies explicitly seek to overcome these techniques. For instance, Di et al. [26] propose a game-theoretic framework that incorporates MIAs into the design of unlearning algorithms. Similarly, Hatua et al. [54] leverage the relationship between MU and MIAs to develop a novel approach based on *Generative Adversarial Network* (GAN) models.

Building on MU, a specialized subfield that has recently gained attention is *federated unlearning* (FU), which addresses the unique challenges of data deletion in *federated learning* (FL) environments. In such settings, the right to be forgotten introduces additional technical obstacles due to the decentralized and heterogeneous nature of these systems [77, 131].

In recent years, the popularity of generative AI, especially in the context of text and image generation models, has been steadily increasing. The widespread adoption of such technologies has prompted the consideration of social and ethical risks that extend beyond existing concerns like biases and liability identification. For instance, Knott et al. [71] argue for the necessity of methods "that allow users to query, for an arbitrary item of content, whether the item was generated (wholly or partly) by the model" as a means to address risks related to fraud and deception. The surge in the volume of data used to train these models (for instance, DALL-E was trained with over 400 million images) has made it increasingly challenging to trace their origins, underscoring the importance of considering data deletion in these models. In this context, Kong and Alfred [72] propose a density-ratio-based framework for efficient deletion applied to GANs.

In the context of generative AI, data inference attacks —related to the previously mentioned MIAs— have emerged as a significant concern for data deletion. In essence, research has demonstrated, through the design of various attack strategies [174, 176], that certain models retain sufficient information from the training dataset to allow the recovery of that information at a later stage. If such techniques were to become widespread, it could even be possible to recover data that was meant to be deleted. This represents a significant privacy risk and highlights a key motivation for addressing data deletion more rigorously. Several authors have tackled this issue from different angles. For instance, Galende et al. [43] highlight the lack of theoretical foundations in this area and propose an evaluation metric to assess the problem. In contrast, Zhang et al. [173] have successfully implemented data forgetting in deep models for image retrieval.

Finally, it is important to emphasize that we must consider the impact of data deletion on models' metrics. Weng [162] highlights the "hiding data problem" as one of the two main misconducts related to Deep Learning. This issue, which involves obtaining altered results due to the selective removal of training data, can occur even unintentionally. This becomes particularly relevant when the deletion is related to one of the model objectives.

5.7 Datasets in Artificial Intelligence

As is well known (and as we have just reviewed in the previous discussions), a key element for the development of any AI and ML activity is a correct use of datasets. However, as acknowledged by Gebru et al. [45] "documenting the creation and use of datasets has received even less attention" (referring to the fact that the notion of data provenance, already discussed in this article, has very recently started to be considered in the field of ML [78]). For instance, Tchanjou et al. [100] emphasize the absence of integration of version control platforms to support ML projects. In this context, they propose a framework aimed at enhancing data traceability and versioning (in line with what we have stated in previous sections), thereby providing certain assurances in data deletion. However, this approach may be insufficient, as it does not consider the use of unlearning techniques.

To the best of our knowledge, one of the few articles that specifically addresses the problem of dataset management in ML and its relation with deletion is that of Paullada et al. [112]. In this work we can find several statements that are worrying to say the least: "Even when these datasets are flagged for removal by the creators, researchers will still attempt to make use of that now illicit information through derivative versions and backchannels [...] Another concerning example of data reuse occurs when derivative versions of an original dataset are distributed—beyond the control of its curators—without any actionable recourse for removal." Note that this quote includes the notion of *derivative dataset* (also pointed out by Thylstrup [150] from an ethical point of view), which is obviously related to the concept of deletion (since a version of a dataset will be generated by any of the usual operations of addition, modification or deletion of data). Of course, this notion, and in general everything that has to do with dataset management, is not exclusive to ML, but can affect many other branches of computer science [13, 114].

5.8 Other Technologies

There are other technologies that maintain some specific relationships with the notion of *deletion* that are worth mentioning. First, we can highlight *data mining* and *process mining*. They are technologies with a shared goal, the discovery of *a priori* hidden information patterns, which may have to do with deletion in at least two directions. On the one hand, it may make sense to analyze the results of mining procedures when data sources are known to have been removed (How are mining procedures affected by deletion? Are they still able to detect these hidden patterns?). On the other hand, these mining procedures can also be used when privacy issues need to be addressed. Let us discuss several examples.

Arefi's doctoral thesis [4] uses data mining on Chinese social networks to discover factors evidencing the deletion of posts on these networks (which has not only an ethical reading, described in Section 4, but also a political one in this case). From a much more technical perspective regarding data mining, Wang et al. [159] study the management and maintenance of sequential patterns (discovered by mining) when records have been deleted. Regarding process mining, privacy challenges have been studied in various contexts. For example, in human-centered industrial environments, Mannhardt et al. [84] analyze how process mining approaches are affected by regulations such as GDPR, and propose privacy guidelines in that particular setting. In general, the term *privacy preserving* is frequently used in these fields. Since data and process mining can be used to uncover hidden information (for instance, anonymized information), it is essential to ensure that these technologies are respectful of privacy. Some studies in this regard are that of Toshniwal [151] and Bhandari and Pahwa [12], both with reviews of different privacy-preserving techniques in data mining; Batista et al. [9] who make a specific approach to privacy-preserving process mining; and Pika et al. [116] that describe a privacy-preserving process mining framework in the sensitive context of healthcare.

Within the framework of digital identity verification, the *Verifiable Credentials* (VCs) [122, 144] represent a key technology—particularly due to their support for selective disclosure. Selective disclosure of VCs "allows a holder to provide a verifier with precisely the information they need and nothing more" [144], in line with the principles of proportionality and necessity. In particular, three main approaches for achieving selective disclosure are identified in [122]: hash-based methods (such as, hash lists and Merkle trees [39]), signature-based methods (such as, CL and BBS+ signatures [39]), and Zero-Knowledge Proof (ZKP) methods. Verifiable credentials can use Decentralized Identifiers (DIDs) [135] for expressing identifiers associated with entities and ZKP protocols to ensure tamper-evident presentation of information [144]. By employing ZKP protocols, personal attributes are proven without being revealed, so that confidentiality is preserved and personal data privacy is not compromised [122]. For example, zkFaith protocol [96] safeguards individual identity attributes while ensuring secure authentication. A recent survey [88] provides a detailed technical account of the operational mechanisms of DIDs and VCs, including how selective disclosure is implemented in practice. Specifically, real frameworks such as DIDKit or IOTA Identity implement it using Selective Disclosure JSON Web Token (SD-JWT) and Zero-Knowledge Selective Disclosure (ZKSD). As previously mentioned, to the extent that data is minimized, it is easier for organizations to comply with data deletion regulations.

The last technological area that we will highlight is that of *software dependency management* [111]. It might seem that it is not related to the concept of *data deletion*, but the idea that *code is data* is not new (*metaprogramming*, understood as the type of programming that uses other programs as data, has been studied for more than thirty years [21]), and it is more relevant since the advent of tools like Github Copilot[6] (that has been "trained on billions of lines of code"). From this perspective, *code deletion* can be considered analogous to *data deletion*. The serious repercussions of uncontrolled code deletion, in terms of dependencies, are well known. The case of the removal of the *left-pad* package from the *npm* repository is one of the most notorious, because of its 'cascading' effect on

[6] https://github.com/features/copilot. (Last visited on December, 2025).

the unavailability of numerous software components (since many of them were unaware of the existence of a dependency on *left-pad*) [19]. In general, deletion of code is one of the factors that can aggravate the development of what is known as *dependency hell* [36], and there is a wealth of literature investigating the problem of *code breakage* and how to avoid it (see, for instance, [81]).

6 Discussion

The preceding sections highlight the multifaceted complexity of data deletion, which cuts across regulatory obligations, privacy guarantees, technical infrastructure, and data lifecycle practices. In this section, we present a structured discussion of the results, summarizing core lessons learned, offering key takeaways, and identifying future research directions of open challenges. Both the key takeaways and the corresponding open research challenges are presented in Table 1.

The human side of deletion. A recurring pattern across the literature is the lack of transparency and verifiability in deletion processes, a problem deeply tied to the human factor: it affects not only individuals requesting deletion, who need assurance that their data has been permanently removed, but also those responsible for executing deletions, who often lack clear mechanisms or feedback to ensure their actions are effective [123, 148]. At this respect, Tebernum and Howar [148] distinguish three distinct "who" roles in deletion: the person(s) that should delete the data (here referred to as the *user*), the executing instance responsible for carrying out deletion (here referred to as the *system administrator*), and the person responsible for the data itself —if different from the previous one (here referred to as the *data custodian*).

These last two roles are typically assumed by entities and organizations (such as service providers, IT teams, or cloud vendors) that develop technology or provide services to users. As highlighted by Ramokapane and Rashid [123], these organizations are responsible for complying with data deletion regulations, particularly the 'right to be forgotten' under GDPR [33], which requires them to implement processes and mechanisms to ensure prompt and effective fulfillment of users' data deletion requests. To operationalize such mandates, it is essential, as described in [170], to have a comprehensive framework for auditing and monitoring data deletion and erasure practices. Regular audits help identify gaps in current processes and highlight areas for improvement. These evaluations should examine: the efficiency of deletion tools and techniques; adherence to applicable legal and regulatory requirements; and the overall security of data management systems. By defining clear metrics and performance indicators, organizations can continuously assess and enhance their data deletion strategies [170]. Although metrics are employed in various contexts —for example, in data forensics by measuring the number of data points recovered with specific recovery algorithms [53], or in AI and machine learning by using Explainable AI metrics, such as Heatmap Coverage and Attention Shift [154]— evaluating these strategies remains difficult. To our knowledge, there is no scientific consensus on standardized evaluation methods or benchmarks for comparison.

Users, for their part, are often left in the dark, encountering challenges such as opaque request handling, insufficient feedback about deletion outcomes, and weak assurance that data has actually been erased [90, 123, 124] —a perception that is well supported by empirical research. For example, Ramokapane et al. [124] found that cloud users "consider information on cloud deletion scarce, not useful, and that it is usually presented to them at the wrong time through a wrong channel", and Minaei et al. [90] reported that over 80% of the social media users surveyed have deleted social media content, yet they reported that popular deletion mechanisms "are not very effective in protecting the privacy of those deletions", indicating a strong desire for clearer, more reliable feedback on deletion outcomes.

As highlighted by Ramokapane and Rashid [123], providing transparent deletion mechanisms not only supports legal compliance, but also builds trust and confidence among users, partners,

Table 1. Key Takeaways and Open Challenges

Takeaway	Open Challenge
The human side of deletion	
• Especially from the users' perspective, data deletion is often opaque and unverifiable since they expect visibility and assurance when data is deleted. • Entities and organizations need to have a comprehensive framework for auditing and monitoring data deletion and erasure practices. • There is no consensus on standardized metrics to assess deletion success.	• Exploring the influence or impact of human factors in deleting data. • Implementing mechanisms that allow users to verify and understand how their deletion requests are processed. • Establishing clear contractual agreements with service providers and a precise definition of responsibilities among involved individuals for deletion implementation and monitoring are essential, yet remain unresolved in many settings. • Defining standardized deletion effectiveness metrics.
The illusion of deletion	
• Data deletion, traditionally framed as a terminal operation in the data lifecycle, is increasingly revealed to be anything but final. • Infrastructure components —e.g., backup systems and cloud platforms— should be designed with data deletion in mind, incorporating features that facilitate compliance with deletion requests (both by individuals or by a legal/regulatory body).	• Developing tools and methods that enable efficient, comprehensive, and verifiable data deletion —particularly in cloud infrastructures. • Future systems should include deletion as a main design concern, with support for traceability and verification.
It depends: deletion in context	
• There is no universal standard for data deletion; data deletion technologies must be tailored to the specific context. • Different deletion techniques, such as logical deletion, cryptographic erasure, and assured deletion, have varying appropriateness depending on factors like data sensitivity and legal requirements.	• Balancing diverse factors —such as data sensitivity, compliance requirements, ethical implications, and practical operational needs— presents a significant challenge in selecting effective data deletion strategies.
When context meets complexity	
• Data deletion must be reframed not just as a technical task, but as an ethical and societal responsibility —especially in domains involving human data. • Data deletion rights granted by frameworks like GDPR, CCPA, CPRA, and VCDPA are often based on assumptions that do not align with the realities of distributed, cloud-based, or backup data systems.	• Many organizations still find it challenging to develop cohesive data governance frameworks that adequately support effective and compliant deletion.
Designed to remember, not to forget	
• Infrastructures such as blockchain, temporal databases, and provenance platforms are explicitly designed to retain information —which inherently clashes with the objective of deletion— making retrofitting deletion support inefficient or unreliable. • Machine unlearning is conceptually crucial for rights like GDPR's 'right to be forgotten,' but current unlearning techniques often fall short: they may not fully remove sensitive data traces nor preserve model integrity.	• Systems designs based on blockchain, temporal databases, or provenance must explore deletion-aware architectures to align with evolving privacy and legal requirements. • Need for robust, verifiable unlearning frameworks, clear metrics for verifying true data removal, and assurance that unlearning does not compromise overall model performance or leave privacy vulnerabilities.

and customers —offering clear business benefits. Additionally, by incorporating explainability around deletion procedures and outcomes, organizations can further enhance accountability while empowering users with a more comprehensive understanding of how deletion processes actually function. To meet these expectations, they emphasize the need for *explainable deletion* underscoring that the technical act of deletion is often opaque to users and difficult to verify, and aiming at making these procedures more understandable. However, integrating such deletion explanations into existing systems remains an ongoing and complex challenge [123]. More broadly, future research should investigate how to systematically integrate the human factor into deletion processes [148].

The illusion of deletion. Data deletion, traditionally framed as a conclusive endpoint in the data lifecycle, is increasingly recognized as an incomplete process [170]. As we have shown throughout the article, although diverse methods (ranging from basic logical deletion to full physical destruction) are available, it is recognized that *deleted* data frequently persist in cloud infrastructures, backups, or even within machine learning models. As such, relying purely on assured deletion requires infrastructure guarantees that are rarely met in practice [123]. For example, cloud storage and backup systems present unique challenges for data deletion, particularly concerning the 'right to be forgotten' [119]. Cloud platforms routinely replicate and back up data across multiple locations, to protect against data loss and meet compliance requirements. However, these same practices can impede deletion by creating copies in places not governed by the same lifecycle rules as primary storage. In this context, advanced techniques such as cryptographic erasure (guided by NIST SP 800-88 standards [70]) have proven promising [2] even for regulatory compliance, by rendering data inaccessible through secure deletion of encryption keys. However, they depend on the implementation of fine-grained operational techniques for more precise and efficient erasure (encryption keys must be careful managed and securely destroyed, data must be encrypted consistently, backup and replication policies must support key destruction, etc.) [2, 70, 170].

It depends: deletion in context. Although data deletion encompasses a range of methods, from basic logical deletion, cryptographic erasure, to assured deletion, the appropriateness of a method is highly dependent on contextual factors, including data sensitivity (Personal Identifiable Data such as medical or financial records, etc.), legal and regulatory constraints (GDPR [33], CCPA [60], CPRA [149]), etc.), ethical concerns (post-mortem data management), ecological considerations (environmental impact, energy use, storage costs), scientific reproducibility needs (validating research using historical data), educational policies (student data retention vs. anonymization). Thus, deletion methods should be carefully chosen based on the context. For instance, in administrative systems, institutions may rely on logical deletion to maintain audit trails —for example, marking records as "inactive" rather than removing them— to support accountability and reporting needs. Additionally, Personal Identifiable Data typically demands sanitization methods (non-recoverable), such as those guided by NIST SP 800-88 standards [70] —using *clear* for overwriting or *purge* for cryptographic erasure, or *destroy* for physical destruction (when dealing, for example, with highly sensitive data). Moreover, as advanced, cryptographic erasure is also commonly employed in cloud platforms, as an efficient method for large-scale deletion.

When context meets complexity. All these contextual factors generate tensions across technological systems and the human parties involved. That is, every constraint and requirement —whether legal, ethical, environmental, or technical— introduces conflicts that affect both the design of data deletion mechanisms and the responsibilities of those who request, implement, or oversee these processes (users, system administrators, custodians). For example, as we have shown, legal frameworks such as the GDPR [33], CCPA [60], CPRA [149] or VCDPA [156], grant individuals strong rights to request data deletion. However, these rights often assume a simplistic model of data control and deletion, which does not reflect the complexities of modern data infrastructures (such as distributed systems, cloud environments, or backups). As highlighted in [170], case studies offer valuable insights into the operational difficulties of implementing deletion in practice. In one notable instance, a leading technology company incurred significant penalties due to its inability to comply with data deletion requests. Upon examination, it was found that the company had an inadequate data governance framework, which had led to fragmented data storage systems and ineffective deletion processes. This case reinforces the importance of implementing a comprehensive data governance strategy that integrates data deletion and erasure as core components [170]. Another example arises in the ecological context, and is related to “dark data” —a notion previously introduced— that is, information and data that businesses collect but usually no longer actively use or analyze. According to Capacity Media [91], the environmental impact of this unused data is staggering, with around 6.4 million tonnes of CO_2 emitted annually into the atmosphere every year (equivalent to the carbon footprint of 80 countries) simply to store it. This creates a clear tension between the human drive to reduce carbon footprint and the technical complexity of implementing reliable deletion mechanisms in modern infrastructures.

Designed to remember, not to forget. Delete processes are in tension with core requirements such as reproducibility, accountability, or traceability. Many systems such as ledgers, temporal databases, or provenance platforms, retain or preserve data by default, which directly conflicts with deletion principles that aim at eliminating data. In the particular case of provenance-aware systems, several approaches presented in the literature, such as UML2PROV [132] or PASS [94] (see also a survey of such systems at [115]), have been designed to automatically capture provenance data on applications. As a result, deletion is especially challenging in provenance-aware systems, which are inherently built to record and preserve metadata rather than to support its removal. These systems rarely include native support for purging provenance data once captured, which inherently conflicts with deletion goals. Similarly, while the blockchain technology provides strong guarantees of traceability and integrity, its core immutability makes deletion fundamentally incompatible. Attempts to reconcile this —such as *redactable blockchains*— while showing potential, have yet to see widespread adoption [143]. Perhaps in these cases (ledgers, temporal databases, or provenance platforms), deletion would need to be architecturally embedded and not layered post hoc. Regarding AI systems, particularly those based on Machine Learning, deletion is not only about removing data but also erasing its influence on trained models. Approaches such as *machine unlearning* seek to achieve this effect; however, current unlearning techniques may fail to guarantee complete removal of sensitive data or even compromise model integrity, which poses fundamental threats to the security and reliability of this technique [106, 167]. This underscores the need for reliable verification methods to boost the credibility of machine unlearning; yet, verification frameworks for unlearning remain fragmented and underdeveloped [106, 167, 171], lacking unified definitions, systematic evaluation protocols, and robust mechanisms to validate whether unlearning has truly occurred [167].

All in all, the highlighted aspects reveal that data deletion cannot be treated as a terminal technical operation. Instead, deletion should be understood as an ongoing, core process, that aligns with various considerations, including ethical standards, legal frameworks, or ecological impacts, and integrates with technical aspects such as data provenance, blockchain, and AI system requirements. All of this supports the conclusion that deletion should be treated as a first-class concern, embedded fundamentally in system architecture, user interactions, legal compliance, and lifecycle processes.

7 Conclusions

The need to delete data stored on computers and other devices is beyond any doubt. In this article we have discussed this topic by means of a three-dimensional analysis. First of all, we have reviewed the terminology related with *data deletion* and we have shown that there are very varied terms without a standardized meaning. Second, we have addressed the different approaches that, from several fields of reference, have been carried out on *data deletion*. Especially noteworthy is the regulation on the “right to be forgotten” that has placed *data deletion* beyond the technical realm, establishing it as a key concern from legal, ethical, and scientific perspectives. In addition, we have described the distinct arguments that arise when a deletion decision is tackled from different fields, showing the difficulties in choosing a clear and safe option. Finally, the relationships between different specific computer technologies and the concept of *deletion* have been analyzed.

Through this analysis, we have that *data deletion* encompasses three different dimensions that are far from trivial. Researchers and practitioners should consider these aspects in any data management effort. Specifically, we argue for the need to establish a holistic, structured framework that organizes the various terms, techniques, and perspectives related to data deletion. The conceptual framework of *Data Governance* [1] emerges as a promising candidate to address this challenge, although significant research and development remain necessary to achieve a comprehensive solution.

Our contribution thus provides a comprehensive and integrated view of data deletion, connecting its diverse terminology, application fields, and the underlying technologies shaping data deletion. This holistic perspective allows us to foreground underexplored dimensions within existing summaries —such as ecological concerns, educational contexts, and the tensions between privacy and scientific reproducibility— while also addressing emerging challenges, including those posed by blockchain persistence and AI-driven approaches like machine unlearning. With this work, we aim at reaffirming deletion as a central, interdisciplinary concern for both researchers and practitioners.

References

[1] Rene Abraham, Johannes Schneider, and Jan Vom Brocke. 2019. Data governance: A conceptual framework, structured review, and research agenda. *International Journal of Information Management* 49 (2019), 424–438.

[2] Sultan Ahmad and Mohammad Mazhar Afza. 2018. A review of assured data deletion mechanism in cloud computing. *International Journal of Engineering and Technology* 7, 45 (2018), 329–332. DOI: https://doi.org/10.14419/ijet.v7i4.5.20101

[3] Dlzar Al Kez, Aoife M. Foley, David Laverty, Dylan Furszyfer Del Rio, and Benjamin Sovacool. 2022. Exploring the sustainability challenges facing digitalization and internet data centers. *Journal of Cleaner Production* 371 (2022), 133633.

[4] Meisam Navaki Arefi. 2020. *Data Mining of Chinese Social Networks: Factors That Indicate Post Deletion.* Ph.D. Dissertation. The University of New Mexico. Retrieved from https://www.proquest.com/docview/2695603150

[5] Giuseppe Ateniese, Bernardo Magri, Daniele Venturi, and Ewerton R. Andrade. 2017. Redactable blockchain - or - rewriting history in bitcoin and friends. In *Proceedings of the 2017 IEEE European Symposium on Security and Privacy.* IEEE, Paris, France, 111–126. DOI: https://doi.org/10.1109/EuroSP.2017.37

[6] Manos Athanassoulis, Subhadeep Sarkar, Tarikul Islam Papon, Zichen Zhu, and Dimitris Staratzis. 2022. Building deletion-compliant data systems. *IEEE Data Engineering Bulletin* 45, 1 (2022), 21–36. Retrieved from http://sites.computer.org/debull/A22mar/p21.pdf

[7] Sharen Bakke, Robert H. Faley, Alan A. Brandyberry, and Marvin D. Troutt. 2005. The impact of privacy concerns on the use of information technologies: A preliminary conceptual model. In *Proceedings of the A Conference on a Human Scale. 11th Americas Conference on Information Systems.* Deepak Khazanchi and Ilze Zigurs (Eds.), Association for Information Systems, Omaha, Nebraska, USA, 209. Retrieved from http://aisel.aisnet.org/amcis2005/209

[8] Konstantia Barmpatsalou, Tiago Cruz, Edmundo Monteiro, and Paulo Simoes. 2018. Current and future trends in mobile device forensics: A survey. *ACM Computing Surveys* 51, 3 (2018), 31 pages. DOI: https://doi.org/10.1145/3177847

[9] Edgar Batista, Antoni Martínez-Ballesté, and Agusti Solanas. 2022. Privacy-preserving process mining: A microaggregation-based approach. *Journal of Information Security and Applications* 68 (2022), 103235. DOI: https://doi.org/10.1016/j.jisa.2022.103235

[10] Aurelie Bayle, Mirko Koscina, David Manset, and Octavio Perez-Kempner. 2018. When blockchain meets the right to be forgotten: Technology versus law in the healthcare industry. In *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence.* IEEE Computer Society, Santiago, Chile, 788–792. DOI: https://doi.org/10.1109/WI.2018.00133

[11] Alan Mark Berg, Stefan T. Mol, Gábor Kismihók, and Niall Sclater. 2016. The role of a reference synthetic data generator within the field of learning analytics. *Journal of Learning Analytics* 3, 1 (2016), 107–128.

[12] Neetika Bhandari and Payal Pahwa. 2019. Comparative analysis of privacy-preserving data mining techniques. In *Proceedings of the International Conference on Innovative Computing and Communications.* Siddhartha Bhattacharyya, Aboul Ella Hassanien, Deepak Gupta, Ashish Khanna, and Indrajit Pan (Eds.), Springer Singapore, Singapore, 535–541.

[13] Anant P. Bhardwaj, Souvik Bhattacherjee, Amit Chavan, Amol Deshpande, Aaron J. Elmore, Samuel Madden, and Aditya G. Parameswaran. 2015. DataHub: Collaborative data science and dataset version management at scale. In *Proceedings of the 7th Biennial Conference on Innovative Data Systems Research.* www.cidrdb.org, Asilomar, CA, USA, 1–7. Retrieved from http://cidrdb.org/cidr2015/Papers/CIDR15_Paper18.pdf

[14] BitRaser. 2021. *Unveil Data Destruction*. BitRaser, Houston, United States. Retrieved from https://www.bitraser.com/ebook. Last visited on December, 2025.

[15] Denis Butin, Denise Demirel, and Johannes Buchmann. 2016. Formal policy-based provenance audit. In *Proceedings of the Advances in Information and Computer Security*. Kazuto Ogawa and Katsunari Yoshioka (Eds.), Springer International Publishing, Cham, 234–253.

[16] Zihao Cao, Jianzong Wang, Shijing Si, Zhangcheng Huang, and Jing Xiao. 2022. Machine Unlearning Method Based On Projection Residual. (2022). arXiv:2209.15276. Retrieved from https://arxiv.org/abs/2209.15276

[17] Andreea Chebac. 2023. What Is Data Erasure? (2023). Retrieved December, 2025 from https://heimdalsecurity.com/blog/what-is-data-erasure/

[18] C. Chen, H. Liu, H. Chi, and P. Xiong. 2025. Enhancing privacy in machine unlearning: Posterior perturbation against membership inference attack. *International Conference on Algorithms and Architectures for Parallel Processing* (2025), 231–248. DOI: https://doi.org/10.1007/978-981-96-1551-3_16

[19] Atique Reza Chowdhury, Rabe Abdalkareem, Emad Shihab, and Bram Adams. 2022. On the untriviality of trivial packages: An empirical study of npm javascript packages. *IEEE Transactions on Software Engineering* 48, 8 (2022), 2695–2708. DOI: https://doi.org/10.1109/TSE.2021.3068901

[20] Peter Christen. 2018. *Data Scrubbing*. Springer New York, New York, NY, 844–848. DOI: https://doi.org/10.1007/978-1-4614-8265-9_80621

[21] James R. Cordy and Medha Shukla. 1992. Practical metaprogramming. In *Proceedings of the 1992 Conference of the Centre for Advanced Studies on Collaborative Research*. John E. Botsford, Arthur G. Ryman, Jacob Slonim, and David J. Taylor (Eds.), IBM, Toronto, Ontario, Canada, 215–224. Retrieved from https://dl.acm.org/citation.cfm?id=962215

[22] Cybersecurity and Privacy Standards Committee. 2022. IEEE Standard for Sanitizing Storage, 2883. (2022). Retrieved from https://standards.ieee.org/ieee/2883/10277/. Last visited on December, 2025.

[23] Tobias Dam, Maximilian Henzl, and Lukas Daniel Klausner. 2021. Delete my account: Impact of data deletion on machine learning classifiers. In *Proceedings of the 2021 International Conference on Software Security and Assurance* . IEEE, Altoona, PA, USA, 7–20. DOI: https://doi.org/10.1109/ICSSA53632.2021.00010

[24] Cécile De Terwangne. 2012. Internet privacy and the right to be forgotten/right to oblivion. *Revista de Internet, Derecho y Politica* 1, 13 (2012), 31–43.

[25] Departament of Defense - USA. 2006. DoD 5220.22-M, National Industry Security Program. (2006). Retrieved December, 2025 from https://www.navsea.navy.mil/Portals/103/Documents/TeamShips/SEA21/InactiveShips/Dismantling/DOD_5220.22_M.pdf

[26] Zonglin Di, Sixie Yu, Yevgeniy Vorobeychik, and Yang Liu. 2024. Adversarial Machine Unlearning. (June 2024). arXiv:cs/2406.07687

[27] Sarah M. Diesburg and An-I Andy Wang. 2010. A survey of confidential data storage and deletion methods. *ACM Computing Surveys* 43, 1 (2010), 2:1–2:37. DOI: https://doi.org/10.1145/1824795.1824797

[28] Eladio Domínguez, Beatriz Pérez, Ángel Luis Rubio, María Antonia Zapata, Alberto Allué, and Antonio López. 2017. Developing provenance-aware query systems: An occurrence-centric approach. *Knowledge and Information Systems* 50, 2 (2017), 661–688. DOI: https://doi.org/10.1007/s10115-016-0950-z

[29] Eladio Domínguez, Beatriz Pérez, Ángel Luis Rubio, María Antonia Zapata, Juan Lavilla, and Alberto Allué. 2014. Occurrence-oriented design strategy for developing business process monitoring systems. *IEEE Transactions on Knowledge and Data Engineering* 26, 7 (2014), 1749–1762. DOI: https://doi.org/10.1109/TKDE.2013.166

[30] Z. Doughan and S. Itani. 2024. Machine unlearning, a comparative analysis. *Communications in Computer and Information Science* 2141 CCIS (2024), 558–568. DOI: https://doi.org/10.1007/978-3-031-62495-7_42

[31] Dylan Thomas Doyle, Casey Paul, and Jed R Brubaker. 2025. Assessing support for mortality: An environmental scan of online platforms. *Proceedings of the ACM on Human-Computer Interaction* 9, 2 (2025), 1–18.

[32] Fajar J Ekaputra, Andreas Ekelhart, Rudolf Mayer, Tomasz Miksa, Tanja Šarčević, Sotirios Tsepelakis, and Laura Waltersdorfer. 2021. Semantic-enabled architecture for auditable privacy-preserving data analysis. *Semantic Web* 1, Preprint (2021), 1–34.

[33] European Parliament, Council of the European Union. 2016. Regulation (EU) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *Official Journal of the European Union* L 119 (2016), 1–88. Retrieved from https://eur-lex.europa.eu/eli/reg/2016/679/oj

[34] European Parliament, Council of the European Union. 2022. Decision (EU) 2022/2481 of the european parliament and of the council of 14 december 2022 establishing the digital decade policy programme 2030. *Official Journal of the European Union* L 323 (2022), 4–26. Retrieved from https://eur-lex.europa.eu/eli/dec/2022/2481/oj

[35] European Parliament, Council of the European Union. 2024. Regulation (EU) 2024/1183 of the european parliament and of the council of 11 April 2024 amending regulation (EU) No 910/2014 as regards establishing the european digital identity framework. *Official Journal of the European Union* L 1183 (2024), 1–56. Retrieved from https://eur-lex.europa.eu/eli/reg/2024/1183/oj

[36] Gang Fan, Chengpeng Wang, Rongxin Wu, Xiao Xiao, Qingkai Shi, and Charles Zhang. 2020. Escaping dependency hell: Finding build dependency errors with the unified dependency graph. In *Proceedings of the 29th ACM SIGSOFT International Symposium on Software Testing and Analysis, Virtual Event.* Sarfraz Khurshid and Corina S. Pasareanu (Eds.), ACM, USA, 463–474. DOI: https://doi.org/10.1145/3395363.3397388

[37] Chunhui Feng, Dawei Wu, Tianle Wu, and Lifang Wei. 2024. An MSDCNN-LSTM framework for video frame deletion forensics. *Multimedia Tools and Applications* 83, 29 (2024), 72745–72764.

[38] Chunhui Feng, Yongxiang Zhong, Yigong Huang, and Xiaolong Liu. 2025. TMTC: Trusted multi-modal transformer classification framework for video frame deletion detection. *The Journal of Supercomputing* 81, 7 (2025), 1–25.

[39] Andrea Flamini, Giada Sciarretta, Mario Scuro, Amir Sharif, Alessandro Tomasi, and Silvio Ranise. 2024. On cryptographic mechanisms for the selective disclosure of verifiable credentials. *Journal of Information Security and Applications* 83, C (Jun 2024). https://doi.org/10.1016/j.jisa.2024.103789

[40] Luciano Floridi. 2012. Big data and their epistemological challenge. *Philosophy and Technology* 25 (2012), 435–437.

[41] Eduard Fosch-Villaronga, Peter Kieseberg, and Tiffany Li. 2018. Humans forget, machines remember: Artificial intelligence and the right to be forgotten. *Computer Law and Security Review* 34, 2 (2018), 304–313. DOI: https://doi.org/10.1016/j.clsr.2017.08.007

[42] Katie Z. Gach and Jed R. Brubaker. 2021. Getting your Facebook affairs in order: User expectations in post-mortem profile management. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–29.

[43] B.A. Galende, P.A. Apellániz, J. Parras, S. Zazo, and S. Uribe. 2025. Membership inference attacks and differential privacy: A study within the context of generative models. *IEEE Open Journal of the Computer Society* 6 (2025), 801–811. DOI: https://doi.org/10.1109/OJCS.2025.3572244

[44] Sanjam Garg, Shafi Goldwasser, and Prashant Nalini Vasudevan. 2020. Formalizing data deletion in the context of the right to be forgotten. In *Proceedings of the Advances in Cryptology - EUROCRYPT 2020 - 39th Annual International Conference on the Theory and Applications of Cryptographic Techniques.* Anne Canteaut and Yuval Ishai (Eds.), Lecture Notes in Computer Science, Vol. 12106, Springer, Zagreb, Croatia, 373–402. DOI: https://doi.org/10.1007/978-3-030-45724-2_{1}{3}

[45] Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé Iii, and Kate Crawford. 2021. Datasheets for datasets. *Communications of the ACM* 64, 12 (2021), 86–92.

[46] A.Shaji George and Sabina Mamedova. 2024. Digital afterlife: Preserving online legacies beyond death. *Partners Universal International Innovation Journal* 02, 01 (2024), 1–14. DOI: https://doi.org/10.5281/zenodo.10581860

[47] Antonio Ginart, Melody Y. Guan, Gregory Valiant, and James Zou. 2019. Making AI forget you: Data deletion in machine learning. In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada.* Hanna M. Wallach, Hugo Larochelle, Alina Beygelzimer, Florence d'Alché-Buc, Emily B. Fox, and Roman Garnett (Eds.), Curran Associates Inc., Red Hook, NY, USA, 3513–3526.

[48] Sergiy Gnatyuk, Vasyl Kinzeryavyy, Tetyana Sapozhnik, Iryna Sopilko, Nurgul Seilova, and Anatoliy Hrytsak. 2020. Modern method and software tool for guaranteed data deletion in advanced big data systems. In *Proceedings of the Advances in Artificial Systems for Medicine and Education II.* Zhengbing Hu, Sergey V. Petoukhov, and Matthew He (Eds.), Springer International Publishing, Cham, 581–590.

[49] Steven N. Goodman, Daniele Fanelli, and John PA Ioannidis. 2016. What does research reproducibility mean? *Science Translational Medicine* 8, 341 (2016), 341ps12–341ps12.

[50] Google Inc. 2023. Data deletion on Google Cloud. (2023). Retrieved December, 2025 from https://cloud.google.com/docs/security/deletion

[51] Google Inc. 2023. Google's statement on deletion and retention. (2023). https://policies.google.com/technologies/retention

[52] Marcel Gregoriadis, Robert Muth, and Martin Florian. 2022. Analysis of arbitrary content on blockchain-based systems using bigquery. In *Companion Proceedings of the Web Conference 2022.* ACM, Virtual event, 478–487.

[53] N.A. Harolanuar and N.H. Ab Rahman. 2024. Forensic evaluation of gutmann and dod secure data deletion algorithms. In *Proceedings of the 1st International Conference on Cyber Security and Computing 2024.* IEEE, Melaka, Malaysia, 88–93. DOI: https://doi.org/10.1109/CyberComp60759.2024.10913756

[54] A. Hatua, T. Nguyen, and A.H. Sung. 2024. Machine unlearning using a multi-GAN based model. In *Proceedings of the AIP Conference.* AIP Publishing, Melville, NY, USA, 1–8. DOI: https://doi.org/10.1063/5.0234688

[55] Helene Hellmich and Jesse David Dinneen. 2022. Making space for the future: The importance of deletion for librarianship and information science and the information society. *Information Research* 27, Special issue (2022), 1–9.

[56] H. Hu, Z. Salcic, L. Sun, G. Dobbie, P.S. Yu, and X. Zhang. 2023. Membership inference attacks on machine learning: A survey. *Computing Surveys* 54, 11s (2023), 37. DOI: https://doi.org/10.1145/3523273

[57] Fang Huang. 2022. *Data Cleansing.* Springer International Publishing, Cham, 275–279. DOI: https://doi.org/10.1007/978-3-319-32010-6_300

[58] Silu Huang, Liqi Xu, Jialin Liu, Aaron J. Elmore, and Aditya Parameswaran. 2017. ORPHEUSDB: Bolt-on versioning for relational databases. *Proceedings of the VLDB Endowment* 10, 10 (2017), 1130–1141.

[59] Ihab F. Ilyas and Xu Chu. 2019. *Data Cleaning*. ACM, New York, NY, USA. DOI: https://doi.org/10.1145/3310205

[60] California Legislative Information. 2018. California Consumer Privacy Act. Assembly Bill No. 375 Chapter 55. (2018). Retrieved December, 2025 from https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180AB375

[61] International Data Sanitization Consortium. 2023. Data Sanitization Terminology and Definitions. (2023). Retrieved December, 2025 from https://www.datasanitization.org/data-sanitization-terminology/

[62] Zachary Izzo, Mary Anne Smart, Kamalika Chaudhuri, and James Zou. 2021. Approximate data deletion from machine learning models. In *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics*. Arindam Banerjee and Kenji Fukumizu (Eds.), Vol. 130, PMLR, Virtual Event, 2008–2016. Retrieved from http://proceedings.mlr.press/v130/izzo21a.html

[63] Nesrine Kaaniche, Maryline Laurent, and Sana Belguith. 2020. Privacy enhancing technologies for solving the privacy-personalization paradox: Taxonomy and survey. *Journal of Network and Computer Applications* 171 (2020), 102807.

[64] S. Kalaiselvi and S. Yoga. 2018. A survey for an efficient secure guarantee in network flow. *International Research Journal of Engineering and Technology* 5, 3 (2018), 1423–1425. Retrieved from https://www.irjet.net/archives/V5/i3/IRJET-V5I3320.pdf

[65] Neeti Kapoor, Pradnya Sulke, and Ashish Badiye. 2021. E-waste forensics: An overview. *Forensic Science International: Animals and Environments* 1 (2021), 100034.

[66] Nickson M. Karie and Hein S. Venter. 2015. Taxonomy of challenges for digital forensics. *Journal of Forensic Sciences* 60, 4 (2015), 885–893. DOI: https://doi.org/10.1111/1556-4029.12809

[67] Lukas Karnowski, Maximilian E. Schüle, Alfons Kemper, and Thomas Neumann. 2021. Umbra as a Time Machine. BTW 2021. Gesellschaft für Informatik, Bonn. DOI: 10.18420/btw2021-06

[68] Victor R. Kebande and Ali Ismail Awad. 2024. Industrial internet of things ecosystems security and digital forensics: Achievements, open challenges, and future directions. *ACM Computing Survey* 56, 5 (2024), 37 pages. DOI: https://doi.org/10.1145/3635030

[69] Yassine El Khanboubi, Mostafa Hanoune, and Mohamed El Ghazouani. 2021. A new data deletion scheme for a blockchain-based de-duplication system in the cloud. *International Journal of Communication Networks and Information Security* 13, 2 (2021), 331–339. DOI: https://doi.org/10.17762/ijcnis.v13i2.4975

[70] Richard Kissel, Andrew Regenscheid, Matthew Scholl, and Kevin Stine. 2014. NIST Special Publication 800-88 Rev.1 Guidelines for Media Sanitization. (2014). Retrieved December, 2025 from http://dx.doi.org/10.6028/NIST.SP.800-88r1

[71] Alistair Knott, Dino Pedreschi, Raja Chatila, Tapabrata Chakraborti, Susan Leavy, Ricardo Baeza-Yates, David Eyers, Andrew Trotman, Paul D. Teal, Przemyslaw Biecek, Stuart Russell, and Yoshua Bengio. 2023. Generative AI models should include detection mechanisms as a condition for public release. *Ethics and Information Technology* 25, 55 (2023), 1572–8439. DOI: https://doi.org/10.1007/s10676-023-09728-4

[72] Zhifeng Kong and Scott Alfeld. 2022. Approximate Data Deletion in Generative Models. (2022). arXiv:cs.LG/2206.14439. Retrieved from https://arxiv.org/abs/2206.14439

[73] Stefan Kulk and Frederik Zuiderveen Borgesius. 2014. Google spain v. gonzález: Did the court forget about freedom of expression?: Case C-131/12 google spain sl and google inc. v. agencia española de protección de datos and mario costeja gonzález. *European Journal of Risk Regulation* 5, 3 (2014), 389–398.

[74] Ben Lenard, Alexander Rasin, Nick Scope, and James Wagner. 2021. What is lurking in your backups?. In *Proceedings of the 36th IFIP International Conference on ICT Systems Security and Privacy Protection (IFIP Advances in Information and Communication Technology)*. Audun Jøsang, Lynn Futcher, and Janne Merete Hagen (Eds.), Vol. 625, Springer, Oslo, Norway, 401–415. DOI: https://doi.org/10.1007/978-3-030-78120-0_26

[75] Ming Di Leom, Kim-Kwang Raymond Choo, and Ray Hunt. 2016. Remote wiping and secure deletion on mobile devices: A review. *Journal of Forensic Sciences* 61, 6 (2016), 1473–1492.

[76] Chen Liu, Hoda Aghaei Khouzani, and Chengmo Yang. 2017. Erasucrypto: A light-weight secure data deletion scheme for solid state drives. *Proceedings on Privacy Enhancing Technologies* 1 (2017), 132–148.

[77] Z. Liu, Y. Jiang, J. Shen, M. Peng, K.-Y. Lam, X. Yuan, and X. Liu. 2024. A survey on federated unlearning: Challenges, methods, and future directions. *Computing Surveys* 57, 1 (2024), 38. DOI: https://doi.org/10.1145/3679014

[78] Shayne Longpre, Robert Mahari, Anthony Chen, Naana Obeng-Marnu, Damien Sileo, William Brannon, Niklas Muennighoff, Nathan Khazam, Jad Kabbara, Kartik Perisetla, Xinyi Wu, Enrico Shippole, Kurt Bollacker, Tongshuang Wu, Luis Villa, Sandy Pentland, and Sara Hooker. 2023. The Data Provenance Initiative: A Large Scale Audit of Dataset Licensing and Attribution in AI. (2023). arXiv:cs.CL/2310.16787. Retrieved from https://arxiv.org/abs/2310.16787

[79] Z. Lu, H. Liang, M. Zhao, Q. Lv, T. Liang, and Y. Wang. 2022. Label-only membership inference attacks on machine unlearning without dependence of posteriors. *International Journal of Intelligent Systems* 37, 11 (2022), 9424–9441. DOI: https://doi.org/10.1002/int.23000

[80] Z. Lu, Y. Wang, Q. Lv, M. Zhao, and T. Liang. 2022. FP2 -MIA: A membership inference attack free of posterior probability in machine unlearning. *In Proceedings of the International Conference on Provable Security.* 167–175. DOI: https://doi.org/10.1007/978-3-031-20917-8_12

[81] Christian Macho, Shane McIntosh, and Martin Pinzger. 2018. Automatically repairing dependency-related build breakage. In *Proceedings of the 25th International Conference on Software Analysis, Evolution and Reengineering*. Rocco Oliveto, Massimiliano Di Penta, and David C. Shepherd (Eds.), IEEE Computer Society, Campobasso, Italy, 106–117. DOI: https://doi.org/10.1109/SANER.2018.8330201

[82] Bernd Malle, Peter Kieseberg, Edgar R. Weippl, and Andreas Holzinger. 2016. The right to be forgotten: Towards machine learning on perturbed knowledge bases. In *Proceedings of the International Conference on Availability, Reliability, and Security.* Francesco Buccafurri, Andreas Holzinger, Peter Kieseberg, A Min Tjoa, and Edgar R. Weippl (Eds.), Lecture Notes in Computer Science, Vol. 9817, Springer, Salzburg, Austria, 251–266. DOI: https://doi.org/10.1007/978-3-319-45507-5_{1}{7}

[83] Vincenzo Mangini, Irina Tal, and Arghir-Nicolae Moldovan. 2020. An empirical study on the impact of GDPR and right to be forgotten - organisations and users perspective. In *Proceedings of the 15th International Conference on Availability, Reliability and Security ARES.* Melanie Volkamer and Christian Wressnegger (Eds.), ACM, Virtual Event, 37:1–37:9. DOI: https://doi.org/10.1145/3407023.3407080

[84] Felix Mannhardt, Sobah Abbas Petersen, and Manuel Fradinho Oliveira. 2018. Privacy challenges for process mining in human-centered industrial environments. In *Proceedings of the 14th International Conference on Intelligent Environments.* IEEE, Roma, Italy, 64–71. DOI: https://doi.org/10.1109/IE.2018.00017

[85] Bharat Manral, Gaurav Somani, Kim-Kwang Raymond Choo, Mauro Conti, and Manoj Singh Gaur. 2019. A systematic survey on cloud forensics challenges, solutions, and future directions. *ACM Computing Survey* 52, 6 (2019), 38 pages. DOI: https://doi.org/10.1145/3361216

[86] Marco Anisetti, Annalisa Appice, Claudio Agostino Ardagna, Alessandro Balestrucci, Nicola Bena, Chiara Braghin, Michelangelo Ceci, Ernesto Damiani, Marco De Monte, Nicola Di Mauro, Stefano Ferilli, Corrado Loglisci, Donato Malerba, Francesca Mazzia, Costantino Mele, Paolo Mignone, and Andrea Rizzello Bortone. 2021. *Data Analytics and Ingestion-time Access Control Initial Report.* Technical Report. IMPETUS Horizon 2020 Project. Retrieved from https://www.impetus-project.eu/images/Deliverables/Deliverable_41.pdf. Last visited on December, 2025.

[87] Roman Matzutt, Jens Hiller, Martin Henze, Jan Henrik Ziegeldorf, Dirk Müllmann, Oliver Hohlfeld, and Klaus Wehrle. 2018. A quantitative analysis of the impact of arbitrary blockchain content on bitcoin. In *Proceedings of the International Conference on Financial Cryptography and Data Security.* Springer, New York, NY, USA, 420–438.

[88] Carlo Mazzocca, Abbas Acar, Selcuck Uluagac, Rebecca Montanari, Paolo Bellavista and Mauro Conti, 2025. A survey on decentralized identifiers and verifiable credentials. *IEEE Communications Surveys & Tutorials.* DOI: 10.1109/COMST.2025.3543197

[89] Niall McCrae and Joanna Murray. 2008. When to delete recorded qualitative research data. *Research Ethics* 4, 2 (2008), 76–77.

[90] Mohsen Minaei, Mainack Mondal, and Aniket Kate. 2022. Empirical understanding of deletion privacy: Experiences, expectations, and measures. In *Proceedings of the 31st USENIX Security Symposium.* USENIX Association, Berkeley, CA, USA, 3415–3432.

[91] Melanie Mingas. 2021. Dark data generating CO2 equivalent to that of 80 countries? Capacity Media, 2021.

[92] Luc Moreau and Paolo Missier. 2013. *PROV-DM: The PROV Data Model.* Technical Report. W3C. Retrieved from https://www.w3.org/TR/prov-dm/

[93] Kiran-Kumar Muniswamy-Reddy. 2006. *Deciding How to Store Provenance.* Technical Report. Harvard Computer Science Group Technical Report TR-03-06. Retrieved from http://nrs.harvard.edu/urn-3:HUL.InstRepos:23526114

[94] Kiran Kumar Muniswamy-Reddy, David A. Holland, Uri Braun, and Margo I. Seltzer. 2006. Provenance-aware storage systems. In *Proceedings of the USENIX Annual Technical Conference, General Track.* USENIX Association, Berkeley, CA, USA, 43–56.

[95] Suntherasvaran Murthy, Asmidar Abu Bakar, Fiza Abdul Rahim, and Ramona Ramli. 2019. A comparative study of data anonymization techniques. In *Proceedings of the 5th IEEE International Conference on Big Data Security on Cloud, IEEE International Conference on High Performance and Smart Computing, and IEEE International Conference on Intelligent Data and Security, Washington, DC, USA, May 27-29, 2019.* IEEE, Piscataway, NJ, USA, 306–309. DOI: https://doi.org/10.1109/BigDataSecurity-HPSC-IDS.2019.00063

[96] Mina Namazi, Duncan Ross, Xiaojie Zhu, and Erman Ayday. 2022. zkFaith: Soonami's Zero-Knowledge Identity Protocol. (2022). arXiv:2212.12785. Retrieved from https://arxiv.org/abs/2212.12785

[97] Arvind Narayanan and Vitaly Shmatikov. 2010. Myths and fallacies of 'personally identifiable information'. *Communications of the ACM* 53, 6 (2010), 24–26.

[98] Bonnie Nardi, Bill Tomlinson, Donald J Patterson, Jay Chen, Daniel Pargman, Barath Raghavan, and Birgit Penzenstadler. 2018. Computing within limits. *Communications of the ACM* 61, 10 (2018), 86–93.

[99] Thanh Tam Nguyen, Thanh Trung Huynh, Zhao Ren, Phi Le Nguyen, Alan Wee-Chung Liew, Hongzhi Yin, and Quoc Viet Hung Nguyen. 2024. A Survey of Machine Unlearning. (2024). arXiv:2209.02299. Retrieved from https://arxiv.org/abs/2209.02299

[100] Aquilas Tchanjou Njomou, Alexandra Johanne Bifona Africa, Bram Adams, and Marios Fokaefs. 2021. MSR4ML: Reconstructing artifact traceability in machine learning repositories. In *Proceedings of the 2021 IEEE International Conference on Software Analysis, Evolution and Reengineering*. IEEE, Piscataway, NJ, USA, 536–540. DOI: https://doi.org/10.1109/SANER50967.2021.00061

[101] Martin Hubert Ofner, Kevin Straub, Boris Otto, and Hubert Österle. 2013. Management of the master data lifecycle: A framework for analysis. *Journal of Enterprise Information Management* 26, 4 (2013), 472–491. DOI: https://doi.org/10.1108/JEIM-05-2013-0026

[102] Dong Bin Oh, Kyung Ho Park, and Huy Kang Kim. 2020. De-wipimization: Detection of data wiping traces for investigating NTFS file system. *Computers and Security* 99 (2020), 102034. DOI: https://doi.org/10.1016/j.cose.2020.102034

[103] Carl Öhman and Luciano Floridi. 2018. An ethical framework for the digital afterlife industry. *Nature Human Behaviour* 2, 5 (2018), 318–320.

[104] Miroslav Ölvecký and Darja Gabriska. 2018. Wiping techniques and anti-forensics methods. In *Proceedings of the 16th IEEE International Symposium on Intelligent Systems and Informatics, Subotica, Serbia, September 13-15, 2018*. IEEE, Piscataway, NJ, USA, 127–132. DOI: https://doi.org/10.1109/SISY.2018.8524756

[105] Abdur Rahman Onik, Joseph Brown, Clinton Walker, and Ibrahim Baggili. 2025. A systematic literature review of secure instant messaging applications from a digital forensics perspective. *Computing Surveys* 57, 9 (2025), 1–36.

[106] Muhammed Shafi K. P., Serena Nicolazzo, Antonino Nocera, and Vinod P. 2025. How Secure is Forgetting? Linking Machine Unlearning to Machine Learning Attacks. (2025). arXiv:2503.20257. Retrieved from https://arxiv.org/abs/2503.20257

[107] Harshvardhan J. Pandit. 2020. *Representing Activities Associated with Processing of Personal Data and Consent Using Semantic Web for GDPR Compliance*. Ph.D. Dissertation. Trinity College Dublin, School of Computer Science and Statistics.

[108] Harshvardhan J. Pandit and Dave Lewis. 2017. Modelling provenance for GDPR compliance using linked open data vocabularies. In *Proceedings of the PrivOn@ ISWC*. CEUR-WS.org, Aachen, Germany, 39–40.

[109] Abelardo Pardo and George Siemens. 2014. Ethical and privacy principles for learning analytics. *British Journal of Educational Technology* 45, 3 (2014), 438–450.

[110] Partition Wizard. 2022. What Is Data Sanitization: Wipe vs Erase vs Format vs Delete. (2022). Retrieved December, 2025 from https://www.partitionwizard.com/disk-recovery/data-sanitization.html

[111] Ivan Pashchenko, Duc Ly Vu, and Fabio Massacci. 2020. A qualitative study of dependency management and its security implications. In *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security, Virtual Event, USA, November 9-13, 2020*. Jay Ligatti, Xinming Ou, Jonathan Katz, and Giovanni Vigna (Eds.), ACM, New York, NY, USA, 1513–1531. DOI: https://doi.org/10.1145/3372297.3417232

[112] Amandalynne Paullada, Inioluwa Deborah Raji, Emily M. Bender, Emily Denton, and Alex Hanna. 2021. Data and its (dis)contents: A survey of dataset development and use in machine learning research. *Patterns* 2, 11 (2021), 100336. DOI: https://doi.org/10.1016/j.patter.2021.100336

[113] João Paulo and José Pereira. 2014. A survey and classification of storage deduplication systems. *ACM Computing Surveys* 47, 1 (2014), 30 pages. DOI: https://doi.org/10.1145/2611778

[114] Kenneth Peng, Arunesh Mathur, and Arvind Narayanan. 2021. Mitigating dataset harms requires stewardship: Lessons from 1000 papers. In *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks 1, NeurIPS Datasets and Benchmarks 2021, December 2021, virtual*. Joaquin Vanschoren and Sai-Kit Yeung (Eds.), NeurIPS, Virtual Conference, 1–16.

[115] Beatriz Pérez, Julio Rubio, and Carlos Sáenz-Adán. 2018. A systematic review of provenance systems. *Knowledge and Information Systems* 57, 3 (2018), 495–543. DOI: https://doi.org/10.1007/s10115-018-1164-3

[116] Anastasiia Pika, Moe T. Wynn, Stephanus Budiono, Arthur H. M. ter Hofstede, Wil M. P. van der Aalst, and Hajo A. Reijers. 2020. Privacy-preserving process mining in healthcare. *International Journal of Environmental Research and Public Health* 17, 5 (2020), 1612. DOI: https://doi.org/10.3390/ijerph17051612

[117] Eugenia Politou, Efthimios Alepis, and Constantinos Patsakis. 2018. Forgetting personal data and revoking consent under the GDPR: Challenges and proposed solutions. *Journal of Cybersecurity* 4, 1 (2018), tyy001.

[118] Eugenia Politou, Fran Casino, Efthimios Alepis, and Constantinos Patsakis. 2021. Blockchain mutability: Challenges and proposed solutions. *IEEE Transactions on Emerging Topics in Computing* 9, 4 (2021), 1972–1986. DOI: https://doi.org/10.1109/TETC.2019.2949510

[119] Eugenia A. Politou, Alexandra K. Michota, Efthimios Alepis, Matthias Pocs, and Constantinos Patsakis. 2018. Backups and the right to be forgotten in the GDPR: An uneasy relationship. *Computer Law and Security Review* 34, 6 (2018), 1247–1257. DOI: https://doi.org/10.1016/j.clsr.2018.08.006

[120] Neoklis Polyzotis, Sudip Roy, Steven Euijong Whang, and Martin Zinkevich. 2018. Data lifecycle challenges in production machine learning: A survey. *SIGMOD Rec.* 47, 2 (2018), 17–28. DOI: https://doi.org/10.1145/3299887.3299891

[121] Kumar Rahul and Rohitash Kumar Banyal. 2020. Data life cycle management in big data analytics. *Procedia Computer Science* 173 (2020), 364–371. DOI: https://doi.org/10.1016/j.procs.2020.06.042

[122] Šeila Bećirović Ramić, Ehlimana Cogo, Irfan Prazina, Emir Cogo, Muhamed Turkanović, Razija Turčinhodžić Mulahasanović, and Saša Mrdović. 2024. Selective disclosure in digital credentials: A review. *ICT Express* 10, 4 (2024), 916–934. DOI: https://doi.org/10.1016/j.icte.2024.05.011

[123] Kopo Marvin Ramokapane and Awais Rashid. 2023. ExD: Explainable deletion. In *Proceedings of the 2023 New Security Paradigms Workshop.* Association for Computing Machinery, New York, NY, USA, 34–47. DOI: https://doi.org/10.1145/3633500.3633503

[124] Kopo Marvin Ramokapane, Jose Such, and Awais Rashid. 2022. What users want from cloud deletion and the information they need: A participatory action study. *ACM Transactions on Privacy and Security* 26, 1 (2022), 34 pages. DOI: https://doi.org/10.1145/3546578

[125] Md Raquibuzzaman, Matchima Buddhanoy, Aleksandar Milenkovic, and Biswajit Ray. 2022. Instant data sanitization on multi-level-cell NAND flash memory. In *Proceedings of the 15th ACM International Systems and Storage Conference, Haifa, Israel, June 13 - 15, 2022.* Michal Malka, Hillel Kolodner, Frank Bellosa, and Moshe Gabel (Eds.), ACM, New York, NY, USA, 85–95. DOI: https://doi.org/10.1145/3534056.3534941

[126] Ambrish Rawat, James Requeima, Wessel Bruinsma, and Richard Turner. 2022. Challenges and Pitfalls of Bayesian Unlearning. (2022). arXiv:2207.03227. Retrieved from https://arxiv.org/abs/2207.03227

[127] Meesam Raza, KS Sakila, K Sreekala, and Asmaa Mohamad. 2024. Carbon footprint reduction in cloud computing: Best practices and emerging trends. *International Journal of Cloud Computing and Database Management* 5, 1 (2024), 25–33.

[128] Joel Reardon. 2016. *Secure Data Deletion.* Springer, Cham, Switzerland. DOI: https://doi.org/10.1007/978-3-319-28778-2

[129] Joel Reardon, David A. Basin, and Srdjan Capkun. 2014. On secure data deletion. *IEEE Security and Privacy* 12, 3 (2014), 37–44. DOI: https://doi.org/10.1109/MSP.2013.159

[130] Jenni Reuben, Leonardo A Martucci, Simone Fischer-Hübner, Heather S Packer, Hans Hedbom, and Luc Moreau. 2016. Privacy impact assessment template for provenance. In *Proceedings of the 2016 11th International Conference on Availability, Reliability and Security.* IEEE, Los Alamitos, CA, USA, 653–660.

[131] Nicolò Romandini, Alessio Mora, Carlo Mazzocca, Rebecca Montanari, and Paolo Bellavista. 2025. Federated unlearning: A survey on methods, design guidelines, and evaluation metrics. *IEEE Transactions on Neural Networks and Learning Systems* 36, 7 (2025), 11697–11717. DOI: https://doi.org/10.1109/tnnls.2024.3478334

[132] Carlos Sáenz-Adán, Beatriz Pérez, Francisco J. García Izquierdo, and Luc Moreau. 2022. Integrating provenance capture and UML With UML2PROV: Principles and experience. *IEEE Transactions on Software Engineering* 48, 2 (2022), 53–68. DOI: https://doi.org/10.1109/TSE.2020.2977016

[133] Sebastian Schelter. 2020. Amnesia - a selection of machine learning models that can forget user data very fast. In *Proceedings of the 10th Annual Conference on Innovative Data Systems Research January 12-15, 2020, Amsterdam, Netherlands.* CIDR, Asilomar, CA, USA, 1–9. Retrieved from https://www.cidrdb.org/cidr2020/papers/p32-schelter-cidr20.pdf

[134] Maximilian E Schüle, Josef Schmeißer, Thomas Blum, Alfons Kemper, and Thomas Neumann. 2021. TardisDB: Extending SQL to support versioning. In *Proceedings of the 2021 International Conference on Management of Data.* ACM, New York, NY, USA, 2775–2778.

[135] Johannes Sedlmeir, Reilly Smethurst, Alexander Rieger, and Gilbert Fridgen. 2021. Digital identities and verifiable credentials. *Business and Information Systems Engineering* 63, 5 (2021), 603–613.

[136] Thanveer Shaik, Xiaohui Tao, Haoran Xie, Lin Li, Xiaofeng Zhu, and Qing Li. 2025. Exploring the landscape of machine unlearning: A comprehensive survey and taxonomy. *IEEE Transactions on Neural Networks and Learning Systems* 36, 7 (2025), 11676–11696. DOI: https://doi.org/10.1109/TNNLS.2024.3486109

[137] Supreeth Shastri, Melissa Wasserman, and Vijay Chidambaram. 2019. The seven sins of personal data processing systems under GDPR. In *Proceedings of the 11th USENIX Workshop on Hot Topics in Cloud Computing.* USENIX Association, USA, 1.

[138] Hira Shahzadi Sikandar, Huda Waheed, Sibgha Tahir, Saif U. R. Malik, and Waqas Rafique. 2023. A detailed survey on federated learning attacks and defenses. *Electronics* 12, 2 (2023), 1–18. DOI: https://doi.org/10.3390/electronics12020260

[139] Junsik Sim, Beomjoong Kim, Kiseok Jeon, Moonho Joo, Jihun Lim, Junghee Lee, and Kim-Kwang Raymond Choo. 2023. Technical requirements and approaches in personal data control. *ACM Computing Survey* 55, 9 (2023), 30 pages. DOI: https://doi.org/10.1145/3558766

[140] Janne Skyt, Christian S. Jensen, and Leo Mark. 2003. A foundation for vacuuming temporal databases. *Data and Knowledge Engineering* 44, 1 (2003), 1–29. DOI: https://doi.org/10.1016/s0169-023x(02)00060-5

[141] Sharon Slade and Paul Prinsloo. 2014. Student perspectives on the use of their data: Between intrusion, surveillance and care. In *EDEN Conference Proceedings*. European Distance and E-Learning Network (EDEN), Vienna, Austria, 291–300.

[142] Richard Snodgrass and Ilsoo Ahn. 1986. Temporal databases. *Computer* 19, 09 (1986), 35–42.

[143] Arpish R. Solanki. 2024. Redactable Blockchain Solutions for IoT: A Review of Mechanisms and Applications. (2024). arXiv:cs.CR/2407.05948. Retrieved from https://arxiv.org/abs/2407.05948

[144] Manu Sporny, Dave Longley, David Chadwick, and Ivan Herman. 2025. *Verifiable Credentials Data Model v2.0*. Technical Report. W3C. Retrieved from https://www.w3.org/TR/vc-data-model-2.0/. Last visited on December, 2025.

[145] StartsUpGeek. 2022. Whats the difference between erasure and deleting? (2022). Retrieved December, 2025 from https://www.creative.onl/startupsgeek/data-erasure/

[146] Patrick Stokes. 2015. Deletion as second death: The moral status of digital remains. *Ethics and Information Technology* 17 (2015), 237–248.

[147] Steven Swanson and Michael Wei. 2010. *SAFE: Fast, Verifiable Sanitization for SSDs*. Technical Report. University of California, San Diego. Retrieved from https://cseweb.ucsd.edu/~swanson/papers/TR-cs2011-0963-Safe.pdf. Last visited on December, 2025.

[148] Daniel Tebernum and Falk Howar. 2023. Structuring the end of the data life cycle. In *Proceedings of the 12th International Conference on Data Science, Technology and Applications, Rome, Italy, July 11-13, 2023*. Oleg Gusikhin, Slimane Hammoudi, and Alfredo Cuzzocrea (Eds.), SCITEPRESS, Setúbal, Portugal, 207–218. DOI: https://doi.org/10.5220/0011999300003541

[149] The California Privacy Rights Act of 2020. 2020. (2020). Retrieved December, 2025 from https://thecpra.org/

[150] Nanna Bonde Thylstrup. 2022. The ethics and politics of data sets in the age of machine learning: Deleting traces and encountering remains. *Media, Culture and Society* 44, 4 (2022), 655–671.

[151] Durga Toshniwal. 2018. *Privacy Preserving Data Mining Techniques for Hiding Sensitive Data: A Step Towards Open Data*. Springer, Singapore", 205–212. DOI: https://doi.org/10.1007/978-981-10-7515-5_15

[152] Katherine Tucker, Janice Branson, Maria Dilleen, Sally Hollis, Paul Loughlin, Mark J Nixon, and Zoë Williams. 2016. Protecting patient privacy when sharing patient-level data from clinical trials. *BMC Medical Research Methodology* 16, 1 (2016), 5–14.

[153] Benjamin E. Ujcich, Adam Bates, and William H. Sanders. 2018. A provenance model for the european union general data protection regulation. In *Proceedings of the Provenance and Annotation of Data and Processes*. Khalid Belhajjame, Ashish Gehani, and Pinar Alper (Eds.), Springer International Publishing, Cham, 45–57.

[154] À.P. Vidal, A.S. Johansen, M.N.S. Jahromi, S. Escalera, K. Nasrollahi, and T.B. Moeslund. 2025. Verifying machine unlearning with explainable AI. In *Proceedings of the International Conference on Pattern Recognition*. 458–473. DOI: https://doi.org/10.1007/978-3-031-88223-4_32

[155] Eva A. Vincze. 2016. Challenges in digital forensics. *Police Practice and Research* 17, 2 (2016), 183–194. DOI: https://doi.org/10.1080/15614263.2015.1128163

[156] Virginia Consumer Data Protection Act. 2021. Title 59.1. Trade and Commerce, Chapter 53. (2021). Retrieved December, 2025 from https://law.lis.virginia.gov/vacode/title59.1/chapter53/section59.1-577/

[157] Kerstin N Vokinger, Daniel J Stekhoven, and Michael Krauthammer. 2020. Lost in anonymization—A data anonymization reference classification merging legal and technical considerations. *The Journal of Law, Medicine and Ethics* 48, 1 (2020), 228–231.

[158] W Gregory Voss and Céline Castets-Renard. 2016. Proposal for an international taxonomy on the various forms of the right to be forgotten: A Study on the Convergence of Norms. *Colorado Technology Law Journal* 14, 2 (2016), 281–344.

[159] Ching-Yao Wang, Tzung-Pei Hong, and Shian-Shyong Tseng. 2001. Maintenance of sequential patterns for record deletion. In *Proceedings of the 2001 IEEE International Conference on Data Mining, 29 November - 2 December 2001, San Jose, California, USA*. Nick Cercone, Tsau Young Lin, and Xindong Wu (Eds.), IEEE Computer Society, Los Alamitos, CA, USA, 536–541. DOI: https://doi.org/10.1109/ICDM.2001.989562

[160] W. Wang, C. Zhang, Z. Tian, S. Yu, and Z. Su. 2025. Evaluation of machine unlearning through model difference. *IEEE Transactions on Information Forensics and Security* 20 (2025), 5211–5223. DOI: https://doi.org/10.1109/TIFS.2025.3571666

[161] Sarah Welsh. 2020. Burying the dead (users): A life of data within limits. *AoIR Selected Papers of Internet Research* 2020 (2020), 4. DOI: https://doi.org/10.5210/spir.v2020i0.11360

[162] Juyang Weng. 2023. On "Deep Learning" Misconduct. (2023). arXiv:cs.LG/2211.16350. Retrieved from https://arxiv.org/abs/2211.16350

[163] Vic (J.R.) Winkler. 2011. *Securing the Cloud: Cloud Computer Security Techniques and Tactics*. Elsevier, Amsterdam, Netherlands. DOI: https://doi.org/10.1016/C2009-0-30544-9

[164] Jinbo Xiong, Lei Chen, Md Zakirul Alam Bhuiyan, Chunjie Cao, Minshen Wang, Entao Luo, and Ximeng Liu. 2020. A secure data deletion scheme for IoT devices through key derivation encryption and data analysis. *Future Generation Computer Systems* 111 (2020), 741–753.

[165] Heng Xu, Tianqing Zhu, Lefeng Zhang, Wanlei Zhou, and Philip S. Yu. 2023. Machine unlearning: A survey. *ACM Computing Surveys* 56, 1 (2023), 36. DOI: https://doi.org/10.1145/3603620

[166] Z. Xu, X. Chen, X. Lan, R. Tang, S. Jiang, and C. Shen. 2024. Empowering data owners: An efficient and verifiable scheme for secure data deletion. *Computers and Security* 144 (2024), 103978. DOI: https://doi.org/10.1016/j.cose.2024.103978

[167] Lulu Xue, Shengshan Hu, Wei Lu, Yan Shen, Dongxu Li, Peijin Guo, Ziqi Zhou, Minghui Li, Yanjun Zhang, and Leo Yu Zhang. 2025. Towards Reliable Forgetting: A Survey on Machine Unlearning Verification, Challenges, and Future Directions. (2025). arXiv:cs.LG/2506.15115. Retrieved from https://arxiv.org/abs/2506.15115

[168] Changsong Yang, Xiaofeng Chen, and Yang Xiang. 2018. Blockchain-based publicly verifiable data deletion scheme for cloud storage. *Journal of Network and Computer Applications* 103 (2018), 185–193. DOI: https://doi.org/10.1016/j.jnca.2017.11.011

[169] Li Yang, Teng Wei, Fengwei Zhang, and Jianfeng Ma. 2018. SADUS: Secure data deletion in user space for mobile devices. *Computers and Security* 77 (2018), 612–626.

[170] YouAccel. 2025. Challenges in Data Deletion and Permanent Erasure. (2025). Retrieved December, 2025 from https://youaccel.com/lesson/challenges-in-data-deletion-and-permanent-erasure/premium?srsltid=AfmBOopHmdDEHKlmbjkHxcAoyyZa2nd3pwD

[171] Binchi Zhang, Zihan Chen, Cong Shen, and Jundong Li. 2024. Verification of machine unlearning is fragile. In *Proceedings of the 41st International Conference on Machine Learning.* JMLR.org, Vienna, Austria, Article 2422, 22 pages.

[172] H. Zhang, T. Nakamura, T. Isohara, and K. Sakurai. 2023. A review on machine unlearning. *SN Computer Science* 4, 4 (2023), 13. DOI: https://doi.org/10.1007/s42979-023-01767-4

[173] Peng-Fei Zhang, Guangdong Bai, Zi Huang, and Xin-Shun Xu. 2022. Machine unlearning for image retrieval: A generative scrubbing approach. In *Proceedings of the 30th ACM International Conference on Multimedia.* Association for Computing Machinery, New York, NY, USA, 237–245. DOI: https://doi.org/10.1145/3503161.3548378

[174] Yuheng Zhang, Ruoxi Jia, Hengzhi Pei, Wenxiao Wang, Bo Li, and Dawn Song. 2020. The secret revealer: Generative model-inversion attacks against deep neural networks. In *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, June 13-19, 2020.* Computer Vision Foundation / IEEE, Los Alamitos, CA, USA, 250–258. DOI: https://doi.org/10.1109/CVPR42600.2020.00033

[175] Dong Zheng, Liang Xue, Chao Yu, Yannan Li, and Yong Yu. 2020. Toward assured data deletion in cloud storage. *IEEE Network* 34, 3 (2020), 101–107.

[176] Ligeng Zhu and Song Han. 2020. Deep leakage from gradients. In *Proceedings of the Federated Learning - Privacy and Incentive.* Qiang Yang, Lixin Fan, and Han Yu (Eds.), Lecture Notes in Computer Science, Vol. 12500, Springer, Switzerland, 17–31. DOI: https://doi.org/10.1007/978-3-030-63076-8_2

[177] Jinglin Zou, Debiao He, Sherali Zeadally, Neeraj Kumar, Huaqun Wang, and Kkwang Raymond Choo. 2021. Integrated blockchain and cloud computing systems: A systematic survey, solutions, and challenges. *ACM Computing Surveys* 54, 8 (2021), 36 pages. DOI: https://doi.org/10.1145/3456628

[178] Guy Zyskind, Oz Nathan, and Alex Pentland. 2015. Decentralizing privacy: Using blockchain to protect personal data. In *Proceedings of the 2015 IEEE Symposium on Security and Privacy Workshops, San Jose, CA, USA, May 21-22, 2015.* IEEE Computer Society, Los Alamitos, CA, USA, 180–184. DOI: https://doi.org/10.1109/SPW.2015.27

Received 12 December 2024; revised 23 July 2025; accepted 25 November 2025