

# LLMs Are Not Databases: Memorization, Disclosure, and the Limits of Privacy Law

*Kristian Stout*

ICLE Issue Brief 2026-02-20

## I. Introduction

Debates over the privacy implications of large language models (LLMs) often rest on an intuitive but unexamined premise: that these systems “store” personal data in the same way a database stores records. Regulators and commentators may therefore reason from database-based regulatory frameworks toward LLMs. The analogy has intuitive appeal but obscures both how language models function technically and how privacy risk arises in practice.

This issue brief argues that the database analogy misleads. Empirical research shows that memorization—verbatim or near-verbatim reproduction of training text—is rare relative to modern corpus scale, concentrated in low-entropy and highly duplicated material, and meaningfully mitigated through practices such as data curation, deduplication, and output filtering. LLMs do not maintain retrievable records. Their parameters encode probabilistic relationships across language, rather than tables of stored entries. Treating them as repositories of personally identifiable information risks imposing rules poorly aligned with both technical reality and observed risk.

Much confusion stems from definitions. In empirical machine-learning research, “memorization” refers to an output-observable event: reproduction of training text under specific prompting conditions. In privacy discourse, the term often serves as shorthand for broader concerns, including inference and hallucination. This issue brief uses “memorization” in the narrower empirical sense because it maps most directly onto what law can regulate—evidence of disclosure in outputs. Other risks, such as inaccurate statements about individuals, may raise separate legal questions, but they differ analytically from memorization and should not be treated as proof that a model stores personal data.

Framed this way, the central question is narrow: whether an LLM’s outputs can constitute a standalone privacy violation by disclosing identifiable personal information. That inquiry differs from the legality of training on datasets that include personal data and from a model’s ability to generate plausible statements through statistical inference. The analysis focuses on the model as a deployed system and on its outputs, because existing legal regimes generally attach liability to disclosure, misuse, or failure to safeguard information in defined contexts.

The discussion proceeds in four parts. Section II reviews the technical literature on memorization, explaining how researchers define, measure, and mitigate it and describing its frequency and distribution. Section III distinguishes research-relevant memorization from privacy-relevant harm and explains why hallucination and inference do not themselves demonstrate data leakage. Section IV evaluates how current law addresses these issues, surveying U.S. federal statutes and California’s California Privacy Rights Act to assess regulatory fit. Section V concludes.

Taken together, the evidence supports three propositions. Large language models are not databases, memorization of personal information is atypical, and privacy risk arises primarily at the point of output and use, rather than from internal statistical representations. Regulatory approaches that

disregard these distinctions risk overdeterrence by constraining socially valuable technologies to address marginal or mischaracterized harms.

## II. The Technical and Empirical Reality of LLM Memorization

Whether large language models (“LLMs”) memorize personal data in a way that constitutes a standalone privacy violation requires careful attention to both technology and law. Much of the debate turns on terminology. Engineers, empirical privacy researchers, and legal doctrine use “memorization” differently, which often obscures the relevant question.

For this issue brief, memorization refers to a specific technical event: a model reproduces training text verbatim or near-verbatim, typically under targeted or adversarial prompting. This concept is narrower than the legal category of personal-data processing.<sup>1</sup> LLMs ordinarily generate probabilistic reconstructions from learned token distributions, not stored records, and most outputs do not retrieve identifiable data.

The empirical literature further shows that memorization is conditional, rather than typical. It concentrates in repeated or low-entropy material, arises most often under contrived prompts, and appears at low rates relative to modern training-corpus size. Developers also deploy multiple safeguards—e.g., dataset curation, deduplication, red-teaming, filtering, and decoding controls—to reduce the likelihood of verbatim disclosure, though no technique eliminates the possibility across all prompts.

This section therefore reviews what the evidence demonstrates about memorization and how those findings inform a narrower legal inquiry: whether the models themselves, as distinct from the act of training or deployment, inherently violate privacy law.

### A. What ‘Memorization’ Means in Large Language Models

In modern usage, “memorization” does not mean simply that a model has seen similar material before, nor that an output resembles its training data. Instead, memorization occurs when a model produces the same sequence of tokens—or a sufficiently similar sequence under standard string-distance metrics—that appears in the training data.<sup>2</sup> This definition reflects the dominant approach in empirical research on extraction and leakage. It is not the only one. Other work defines memorization through membership inference, generalization gaps, or differential privacy.<sup>3</sup> For

---

<sup>1</sup> See Regulation (EU) 2016/679 of the European Parliament and of the Council (General Data Protection Regulation) art. 4(2), 2016 O.J. (L 119) 1 (defining “processing” broadly to include “any operation or set of operations” performed on personal data, including collection, storage, use, disclosure, or deletion).

<sup>2</sup> See *Gemma 3 Technical Report*, ARXIV:2503.19786, at 9 (Mar. 25, 2025), <https://arxiv.org/abs/2503.19786> (defining “exact memorization” as token-for-token reproduction and “approximate memorization” as matching within a 10% edit distance); see also *The Llama 3 Herd of Models*, ARXIV:2407.21783 (July 31, 2024), <https://arxiv.org/abs/2407.21783> (discussing tests that probe whether models can reproduce training-data text verbatim).

<sup>3</sup> See Nicholas Carlini et al., *The Secret Sharer: Evaluating and Testing Unintended Memorization in Neural Networks*, in PROCEEDINGS OF THE 28TH USENIX SECURITY SYMPOSIUM (2019), <https://www.usenix.org/conference/usenixsecurity19/presentation/carlini>; Vitaly Feldman & Chiyuan Zhang, *What Neural Networks Memorize and Why: Discovering the Long Tail via Influence Estimation*,

present purposes, we focus on verbatim and near-verbatim reproduction because it most directly relates to output-stage privacy risk and to the legal question of whether a model functions as a repository of personal data.

Under this framework, large language models do not store information as records or tables. Their parameters encode probability distributions over token sequences, not discrete entries comparable to database rows.<sup>4</sup> Most outputs therefore reflect probabilistic reconstruction—novel text generated from learned distributions, rather than retrieved records.

Risk arises when the boundary between reconstruction and retrieval blurs. If a sequence appears frequently in the training data, or occurs in low-entropy boilerplate structures, the model may generate it with unusually high probability under carefully constructed extraction prompts, often continuation-style or otherwise adversarial. Empirical work demonstrates that such leakage is real and measurable.<sup>5</sup>

Defining memorization in practice remains difficult because natural-language documents commonly contain overlapping and repeated text. Material known to be in the training data can closely resemble material that is not, creating an inherently fuzzy boundary between members and non-members.<sup>6</sup> As a result, standard membership definitions may fail to capture relevant leakage. Membership-inference techniques may classify passages that are lexically or semantically similar to members as non-members, even when privacy auditors would still regard the disclosure as meaningful information leakage.<sup>7</sup>

## B. Mitigating Memorization: Techniques and Tradeoffs

Frontier-model developers and the technical literature describe a range of measures designed to reduce verbatim memorization and the risk of disclosing sensitive material. The specific techniques—and the degree of public disclosure—vary across systems.<sup>8</sup>

---

ARXIV:2008.03703 (Aug. 9, 2020), <https://arxiv.org/abs/2008.03703>; Wanrong Zhang et al., *Leakage of Dataset Properties in Multi-Party Machine Learning*, in PROCEEDINGS OF THE 30TH USENIX SECURITY SYMPOSIUM (2021), <https://www.usenix.org/conference/usenixsecurity21/presentation/zhang-wanrong>.

<sup>4</sup> See Michael Duan et al., *Do Membership Inference Attacks Work on Large Language Models?*, ARXIV:2402.07841 (Sept. 16, 2024), <https://arxiv.org/abs/2402.07841> (explaining that the model is an autoregressive language model that predicts the probability distribution over the next token given a prompt).

<sup>5</sup> See Nicholas Carlini et al., *Extracting Training Data from Large Language Models*, in PROCEEDINGS OF THE 30TH USENIX SECURITY SYMPOSIUM (2021), <https://arxiv.org/pdf/2012.07805>; Milad Nasr et al., *Scalable Extraction of Training Data from (Production) Language Models*, ARXIV:2311.17035 (Nov. 28, 2023), <https://arxiv.org/abs/2311.17035>.

<sup>6</sup> Duan et al., *supra* note 4, at 6-7.

<sup>7</sup> *Id.* at 8-9.

<sup>8</sup> See Badrinath Ramakrishnan & Akshaya Balaji, *Assessing and Mitigating Data Memorization Risks in Fine-Tuned Large Language Models*, ARXIV:2508.14062, at 3, 5, 41-42 (Aug. 10, 2025), <https://arxiv.org/abs/2508.14062>; Kunj Joshi et al., *Randomized Masked Finetuning: An Efficient Way to Mitigate Memorization of PII in LLMs*, ARXIV:2512.03310, at 6-7 (Feb. 9, 2026), <https://arxiv.org/abs/2512.03310>.

Reported safeguards begin at training. Developers employ dataset curation and filtering,<sup>9</sup> deduplication to limit repetition-driven regurgitation,<sup>10</sup> and selective inclusion or exclusion of sources more likely to contain personal information. They also deploy post-training and operational controls, including red-teaming, classifier-based filtering, and decoding or sampling heuristics intended to reduce the likelihood that a model emits sensitive or low-entropy strings in response to ordinary prompts.<sup>11</sup> Researchers have also explored differential-privacy-enhanced training, although it generally involves meaningful tradeoffs in performance and utility.<sup>12</sup>

A separate line of research pursues more ambitious interventions, such as machine unlearning or targeted suppression of specific information in a trained model. These techniques remain technically immature.<sup>13</sup> Complex unlearning objectives can degrade model quality or fail to remove information embedded in distributed internal representations.

Mitigation therefore involves tradeoffs. Existing techniques can materially reduce observable memorization and the likelihood of verbatim disclosure, but they do not guarantee elimination of leakage across the full space of prompts, adversarial strategies, and distribution shifts, particularly at frontier scale.<sup>14</sup> More aggressive privacy-preserving approaches remain an active research area and can impose substantial costs in performance, training complexity, and practical utility.

### C. Memorization Is Rare and Conditional

Across multiple studies, verbatim memorization appears uncommon relative to the scale of modern training corpora.<sup>15</sup> Importantly, these estimates do not measure “the fraction of the training set memorized.” They measure the share of test cases in which a model, given a short snippet, continues the passage by reproducing the original training text verbatim. Under that benchmark, even very large models trained on multi-trillion-token datasets show low reproduction rates—on the order of about 1-4% in recent frontier-model evaluations<sup>16</sup>—and newer model generations often copy less than earlier ones.<sup>17</sup>

These tests typically evaluate ordinary text. They do not directly answer a different policy question: whether a model can be induced to output sensitive personal information, which researchers assess using different targets and methods. Against trillion-token corpora, the total amount of text

---

<sup>9</sup> Gemma 3 Technical Report, *supra* note 2, at 2-3.

<sup>10</sup> The Llama 3 Herd of Models, *supra* note 2, at 4-5, 53-54.

<sup>11</sup> The Llama 3 Herd of Models, *supra* note 2, at 47, 49; Gemma 3 Technical Report, *supra* note 2, at 2-3.

<sup>12</sup> See Carlini, *supra* note 5, at 2644.

<sup>13</sup> See A. Feder Cooper *et al.*, *Machine Unlearning Doesn't Do What You Think: Lessons for Generative AI Policy and Research*, ARXIV:2412.06966 (Dec. 9, 2024), <https://arxiv.org/abs/2412.06966v2>.

<sup>14</sup> See Carlini *et al.*, *supra* note 5 (noting some techniques “help mitigate memorization but cannot prevent” it entirely); Da Yu, *Differentially Private Fine-tuning of Language Models*, ARXIV:2110.06500v2 (July 18, 2022), <https://arxiv.org/pdf/2110.06500>.

<sup>15</sup> The Llama 3 Herd of Models, *supra* note 2, at 1.

<sup>16</sup> *Id.* at 41-42; Gemma 3 Technical Report, *supra* note 2, at 9.

<sup>17</sup> *Id.*

extractable verbatim, even with adversarial prompting, remains small. Scale cuts both ways. Larger models may increase copying risk, but improved curation and deduplication can offset it, and newer generations often combine greater scale with cleaner training data.<sup>18</sup>

The strongest extraction methods rely on contrived prompts such as “continue the following sequence exactly,” seeded with partial strings known to appear repeatedly in the corpus.<sup>19</sup> Without this scaffolding, random prompting produces little to no verbatim reproduction.<sup>20</sup>

An extraction study by Nicholas Carlini *et al.* illustrates the point.<sup>21</sup> In more than 600,000 model generations, the researchers confirmed only 604 verbatim extractions—about a 0.1% rate.<sup>22</sup> Even these required highly adversarial sampling and ranking procedures, as well as independent access to the same web text used to train GPT-2 to verify matches.<sup>23</sup> Extraction was therefore possible but rare: confirming memorization required prior access to the underlying data.<sup>24</sup>

A later large-scale analysis reached similar conclusions.<sup>25</sup> Across several major open-source models, only 0.03% to 1.4% of outputs contained recoverable training text, even after massive output generation and the deployment of specialized matching tools.<sup>26</sup> More sophisticated methods did not materially change the result. Extractable memorization exists, but it is a low-base-rate phenomenon.<sup>27</sup> Fine-tuned chat models require additional attack techniques and still produce memorized sequences at rates well below 1% of total output.<sup>28</sup>

Two empirical regularities emerge. First, duplication strongly predicts memorization.<sup>29</sup> Repetition is the most reliable indicator: strings appearing tens, hundreds, or thousands of times—such as license

---

<sup>18</sup> Alexander Xiong *et al.*, *The Landscape of Memorization in LLMs: Mechanisms, Measurement, and Mitigation*, ARXIV:2507.05578, at 2 (Dec. 12, 2025), <https://arxiv.org/abs/2507.05578>.

<sup>19</sup> See *Gemma 3 Technical Report*, *supra* note 2, at 9 (describing memorization tests that use a fixed prompt structure with a 50-token prefix and 50-token suffix to measure exact and approximate memorization).

<sup>20</sup> See *The Llama 3 Herd of Models*, *supra* note 2, (noting that effective memorization audits rely on targeted, structured prompts rather than random inputs); Jamie Hayes *et al.*, *Strong Membership Inference Attacks on Massive Datasets and (Moderately) Large Language Models*, ARXIV:2505.18773v1, at 2 (May 24, 2025), <https://arxiv.org/html/2505.18773v1> (explaining that unstructured prompts rarely elicit verbatim training text and that extraction tests instead target high-frequency sequences likely to be overfit); *id.* (describing sampling prompts and expected outputs based on their frequency in the training corpus); see also *Gemma 3 Technical Report*, *supra* note 2, at 8 (describing contrived prompt structures—*e.g.*, a fixed 50-token prefix used to predict a 50-token suffix—to measure extractable memorization).

<sup>21</sup> See Carlini, *supra* note 5.

<sup>22</sup> *Id.*

<sup>23</sup> *Id.* at 13.

<sup>24</sup> *Id.* at 7.

<sup>25</sup> See Nasr, *supra* note 5.

<sup>26</sup> *Id.* at 7, 19.

<sup>27</sup> *Id.*

<sup>28</sup> *Id.*

<sup>29</sup> *Id.* at 14-15; Duan, *supra* note 4, at 4-5.

blocks, email signatures, news boilerplate, or SEO spam—are far more likely to be reproduced. Deduplicating training corpora can reduce verbatim memorization by an order of magnitude.

Second, low-entropy strings are disproportionately memorized. Highly predictable content—addresses, phone numbers, templated contact lines (“Contact me at ...”), or schematic code—has few plausible continuations and is easier to reproduce verbatim.<sup>30</sup> High-entropy material, such as narrative prose or unique personal messages, rarely reappears in the same form.<sup>31</sup> Memorization risk is therefore unevenly distributed across data types.

Taken together, these patterns indicate that memorization is localized. It does not describe ordinary model behavior. It appears under specific conditions and with specific types of content.

### **III. From Memorization to Disclosure: The Legal Significance of Model Outputs**

The literature contains an important definitional gap. Most machine-learning research uses “memorization” to mean verbatim reproduction of training text. Privacy-law debates often use the term more loosely, implying that a model internally retains personal information as a discrete object. Conflating these concepts obscures the relevant legal inquiry.

From a policy perspective, verbatim reproduction is neither necessary nor sufficient for privacy risk. A model may generate statements about an identifiable person—accurate or not—without reproducing training text verbatim. Conversely, a model may reproduce memorized strings that contain no personal information, such as boilerplate open-source license language.

For this issue brief, the analysis has two distinct components. First, whether personal data can meaningfully be said to reside in a model’s internal parameters. Second, and more legally significant, whether a model emits information in its outputs in a way that constitutes disclosure or misuse.

We focus on the second question. Accordingly, we set aside the permissibility of training on datasets that may contain personal data and instead examine whether a trained model, treated as a discrete object, can itself create a standalone privacy violation through its outputs. As the discussion of U.S. law below shows, liability generally attaches to disclosure, use, or access, not to the mere existence of statistical representations in model weights.

Under this framework, empirical memorization research—centered on verbatim reproduction—is an imperfect proxy for privacy harm. It identifies a measurable category of risk, but it does not capture every way personal information might appear in outputs. Privacy-relevant harm depends not only on

---

<sup>30</sup> See *id.* at 3, 6 (describing high overlap and repetition in domains such as GitHub); *The Llama 3 Herd of Models*, *supra* note 2, at 4, 43 tbl. 24 (discussing filtering of PII and code data); Nils Lukas *et al.*, *Analyzing Leakage of Personally Identifiable Information in Language Models*, ARXIV:2302.00539, at 346 (Feb. 1, 2023), <https://arxiv.org/abs/2302.00539> (examining privacy risks from potential PII leakage).

<sup>31</sup> Duan *et al.*, *supra* note 4, at 8 (noting that repetition and common phrasing are inherent features of natural-language data, making truly unique sequences rare).

memorized sequences, but also on whether information associated with identifiable individuals emerges under prompting, whether through direct regurgitation or transformations, such as paraphrase or translation.

### **A. Personal Data Outputs Are Not Necessarily Memorization**

What does the evidence show about personal data specifically? The empirical record supports several propositions.

Researchers have demonstrated that personal phone numbers, email addresses, and short biographical details can sometimes be extracted from LLMs. These events, however, typically require that the information appears repeatedly online or in templated sources, *e.g.*, faculty directories or scraped social-media biographies. The individuals most at risk therefore tend to have large digital footprints—public figures, academics, journalists, and others with SEO-dense online profiles. By contrast, models are unlikely to reproduce one-off personal information about private individuals, unless the information has been duplicated extensively across the web.

Many outputs that appear to contain personal data reflect probabilistic reconstruction, rather than verbatim retrieval from training data. In deployed systems, outputs that contain apparent personally identifiable information (PII) may also originate from user-provided prompts, uploaded documents, retrieval-augmented generation (RAG), or integrated web search, rather than training-data memorization. Often, the model generates a plausible statement based on statistical associations, rather than stored facts. When a model states that “John Doe is a lawyer in Chicago,” the output may be a hallucination or a statistical interpolation—common name plus common profession plus major city—or a paraphrased reconstruction. In ordinary usage, this does not resemble retrieval of a stored record.

The distinction matters. Treating probabilistic fabrication as equivalent to unauthorized disclosure of factual personal data would impose deterrence disproportionate to the underlying harm, particularly when the asserted personal data was never processed during training.

This leads to a broader point: hallucination is not evidence of memorization. Generative models routinely produce plausible but fabricated claims about individuals based on learned correlations. Equating those outputs with disclosure of training-set personal data collapses two distinct phenomena and treats speculative generation as proof of data leakage.

A regulatory approach that collapses hallucination into leakage would effectively treat statistical inference as the exfiltration of a stored record or the processing of personal data.<sup>32</sup> This concern differs from cases in which outputs combine publicly available information about real individuals

---

<sup>32</sup> Relatedly, if “hallucination” refers to a model’s failure to reproduce accurate information, restricting the use of personal data in training is not an obvious fix. In some settings, adding high-quality, lawfully obtained data (including information about real-world entities) may improve calibration and reduce certain hallucinations, while raising separate privacy issues involving lawful basis, purpose limitation, and data minimization. In short, hallucination risk and training-data privacy risk are distinct; conflating them can lead to data restrictions that do not address the underlying failure mode.

with false or distorted claims, which instead resemble accuracy-based or defamation-style harms.<sup>33</sup> Conflating such hybrid outputs with memorization-driven leakage would distort incentives by encouraging developers to suppress useful generative capabilities to avoid liability for outputs that do not disclose stored records. The likely result would be reduced social utility, including diminished capacity for models to generate content, assist reasoning, or perform tasks that inherently involve interpolation.

Accordingly, the presence of PII in model outputs does not, by itself, show that a model memorized that information in a manner sufficient for a standalone privacy-violation claim. Under U.S. law, privacy liability generally turns on disclosure, unreasonable publicity, breach of confidentiality, or specific statutory triggers—not the existence of statistical associations within a model.

#### **IV. How Existing Privacy Law Applies to Large Language Models**

This section examines how different legal frameworks respond to output-disclosure risks associated with memorization, and how those frameworks map (or fail to map) onto LLM systems that often combine a base model with retrieval, user-supplied inputs, and other data sources.

The foregoing analysis has two immediate implications for policymakers. First, empirical evidence sharply limits the circumstances in which large language models plausibly resemble repositories of personal data. Memorization is rare, context-dependent, and most relevant at the point of output, rather than within internal parameters. Second, existing privacy regimes differ in how closely they track those technical realities.

This section examines how current legal frameworks respond to the memorization question and what that response reveals about regulatory fit. U.S. federal privacy law generally follows a sectoral, actor- and disclosure-based model: liability attaches when regulated entities collect, use, safeguard, or disclose identifiable information in defined relationships and contexts. California’s approach, while broader and more risk-oriented, similarly focuses on business practices, proportionality, and foreseeable harms, rather than the mere existence of statistical representations.

The goal is modest. This issue brief does not propose a comprehensive new privacy regime. Instead, it evaluates whether existing doctrines align with the technical characteristics of LLM systems and identifies where mismatches may produce over- or under-deterrence. The discussion first surveys federal privacy statutes and then turns to California’s hybrid framework, emphasizing how both regimes address output-related disclosure risks in systems that may combine a base model with retrieval tools, user-supplied inputs, and other external data sources.

---

<sup>33</sup> See NOYB – European Center for Digital Rights, *ChatGPT Provides False Information About People, and OpenAI Can’t Correct It* (Apr. 29, 2024), <https://noyb.eu/en/chatgpt-provides-false-information-about-people-and-openai-cant-correct-it>.

### A. Federal Privacy Statutes: Actor- and Disclosure-Based Liability

U.S. federal law generally follows a disclosure- and misuse-based conception of privacy, rather than a comprehensive data-protection model. Liability typically turns on whether a regulated actor disclosed, misused, or failed to safeguard identifiable personal information in defined contexts. None of the major federal statutes treat the transformation of text into statistical parameters as legally relevant storage, memorization, or processing. The legal trigger is misuse of identifiable data, not the existence of internal representations.

More broadly, U.S. privacy law—especially at the federal level—does not treat possession of statistically encoded information as *per se* unlawful. Privacy torts and related statutory provisions ordinarily require conduct such as public disclosure of private facts, intrusion upon seclusion, appropriation of a person’s name or likeness, or false light. Federal statutes including the Health Insurance Portability and Accountability Act (HIPAA), the Fair Credit Reporting Act (FCRA), the Children’s Online Privacy Protection Act (COPPA), the Gramm–Leach–Bliley Act (GLBA), and the Video Privacy Protection Act (VPPA) similarly define specific actors, specific data types, and specific prohibited acts. Across these regimes, liability attaches to identifiable information handled in regulated relationships and communicated to others—not to background exposure during model training.

This structure produces a consistent pattern across the statutes discussed below. Each is sectoral and context-dependent: HIPAA regulates health-care actors, FCRA governs consumer-reporting activities, COPPA addresses child-directed services, GLBA covers financial institutions, and the VPPA targets video-service providers. In each case, obligations arise at collection, use, disclosure, safeguarding, or eligibility decisionmaking. None treats a general-purpose model developer as regulated merely because a model trained on heterogeneous text, and none treats internal model parameters as legally cognizable records.<sup>34</sup>

Accordingly, the legally salient questions concern deployment and outputs. Statutory risk arises when a covered entity uses an LLM in a way that discloses identifiable information, incorporates protected data into eligibility determinations, or fails to implement required safeguards—not when training produces statistical representations that, standing alone, do not identify an individual. This issue brief surveys these frameworks to illustrate how U.S. privacy law evaluates conduct involving information systems and why those rules generally operate at the point of use and disclosure rather than at the level of model architecture.

---

<sup>34</sup> Because, as discussed *infra*, these regimes regulate defined actors and the use or disclosure of identifiable information in specified contexts. Treating model weights as regulated “personal information” would stretch statutory definitions and equate statistical parameters with maintained records. That move would likely draw challenges as inconsistent with statutory text, structure, and historically understood triggers, and as an agency assertion of major new regulatory authority absent clear congressional authorization. See *Loper Bright Enters. v. Raimondo*, 603 U.S. 369 (2024), [https://www.supremecourt.gov/opinions/23pdf/22-451\\_7m58.pdf](https://www.supremecourt.gov/opinions/23pdf/22-451_7m58.pdf).

### *1. HIPAA's Actor-Based Limits*

The Health Insurance Portability and Accountability Act (HIPAA) imposes privacy and security obligations on a defined set of regulated actors: “covered entities,” including health-care providers, health plans, and clearinghouses, as well as their business associates.<sup>35</sup> The statute protects identifiable “protected health information” (PHI),<sup>36</sup> and liability generally arises from the unauthorized use or disclosure of PHI within the health-care ecosystem.<sup>37</sup>

HIPAA does not regulate noncovered entities merely because they process information that contains health-related content. By extension, the statute ordinarily would not apply to statistical or machine-learning models that ingest text containing PHI when developed by a noncovered entity.<sup>38</sup> Absent covered-entity status or a business-associate relationship, a developer building a general-purpose LLM trained on heterogeneous web data is not subject to HIPAA and does not maintain PHI in a legally cognizable form. As a result, HIPAA generally does not reach an LLM’s internal numerical parameters, even if fragments of health-related text contributed to the training corpus.

**Relevance to LLM Memorization:** HIPAA is a sectoral, actor-based regime. It regulates who handles information and imposes duties tied to the use and disclosure of identifiable PHI within that regulated system. That structure makes HIPAA a poor fit for treating a general-purpose LLM’s internal parameters as a repository of PHI. The statute is more plausibly implicated at the output and deployment stage—*e.g.*, when a covered entity or business associate uses an LLM in a way that results in an unauthorized disclosure of identifiable PHI—rather than at the training stage conducted by a noncovered developer.

### *2. FCRA's Role- and Purpose-Based Limits*

The Fair Credit Reporting Act (FCRA) governs a defined set of actors and activities within the consumer-credit ecosystem. It applies primarily to “consumer reporting agencies” (CRAs),<sup>39</sup> defined as entities that, for fees or on a cooperative nonprofit basis, regularly assemble or evaluate consumer information to furnish consumer reports to third parties.<sup>40</sup> A “consumer report” is any communication by a CRA bearing on a consumer’s creditworthiness, character, reputation, personal

---

<sup>35</sup> See Covered Entities and Business Associates, U.S. DEP’T OF HEALTH & HUM. SERVS., <https://www.hhs.gov/hipaa/for-professionals/covered-entities/index.html> (last visited Feb. 17, 2026).

<sup>36</sup> *Id.*

<sup>37</sup> 42 U.S.C. § 1320d-6; *see also* Scope of Criminal Enforcement Under 42 U.S.C. § 1320d-6, U.S. DEP’T OF JUSTICE, OFF. LEGAL COUNSEL (2005), [https://www.justice.gov/sites/default/files/olc/opinions/attachments/2014/11/17/hipaa\\_final.htm](https://www.justice.gov/sites/default/files/olc/opinions/attachments/2014/11/17/hipaa_final.htm).

<sup>38</sup> See Nicolas Terry, *Protecting Patient Privacy in the Age of Big Data*, IND. UNIV. ROBERT H. MCKINNEY SCH. OF L. (Sept. 27, 2012), [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2153269](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2153269), at 21–22 (explaining that health information protected by HIPAA in the hands of a covered entity may become unprotected when obtained or used by a noncovered entity, such as a data-analytics firm).

<sup>39</sup> 15 U.S.C. § 1681(b).

<sup>40</sup> 15 U.S.C. § 1681a(f).

characteristics, or mode of living that is used, or expected to be used, to determine eligibility for credit, insurance, employment, or other statutorily specified purposes.<sup>41</sup>

Regulatory guidance reflects this role-based structure. The Federal Trade Commission (FTC) describes the statute as protecting information collected by credit bureaus, medical-information companies, and tenant-screening services, and emphasizes that such information may be furnished only for permissible purposes.<sup>42</sup> The Consumer Financial Protection Bureau similarly frames FCRA obligations around CRAs and users of consumer reports, including furnishers that supply information to CRAs and entities that rely on reports for eligibility decisions.<sup>43</sup> Academic commentary likewise characterizes FCRA as regulating the collection, maintenance, and disclosure of consumer information by CRAs and related participants in that reporting system.

Liability therefore arises when a CRA fails to meet the statute's accuracy, disclosure, or permissible-purpose requirements, or when a furnisher or user violates duties tied to consumer reports—e.g., by supplying inaccurate data to a CRA or using a report without a permissible purpose. FCRA does not cover every entity that handles consumer information and does not regulate all databases containing personal data. It applies only when three elements align: a qualifying actor, a qualifying communication (a consumer report), and a qualifying use (an eligibility determination or another enumerated purpose).

**Relevance to LLM Memorization:** FCRA is a role- and purpose-based regime, not a general personal-data statute. A general-purpose LLM developer typically does not become a CRA merely because the model trained on heterogeneous text that includes consumer-related information. The statute instead becomes relevant at deployment. If an LLM system generates or supplies information that functions as a consumer report, or is used in an eligibility decision in a way that makes the operator resemble a CRA or covered user, FCRA's accuracy, disclosure, and permissible-purpose requirements may apply. The key question is therefore not whether a model internally "contains" consumer data, but whether the system assembles, evaluates, and communicates consumer information for covered purposes.

### 3. *COPPA's Operator- and Context-Specific Scope*

The Children's Online Privacy Protection Act (COPPA) imposes obligations on a defined group of operators: those that run online services "directed to children" under age 13 or that have actual knowledge they collect personal information from a child.<sup>44</sup> The FTC explains that the law applies

---

<sup>41</sup> 15 U.S.C. § 1681a(d).

<sup>42</sup> 15 U.S.C. §§ 1681–1681x.

<sup>43</sup> Fair Credit Reporting Act, CONSUMER FIN. PROT. BUREAU, Consumer Laws & Regulations (2012), [https://files.consumerfinance.gov/f/documents/102012\\_cfpb\\_fair-credit-reporting-act-fcra\\_procedures.pdf](https://files.consumerfinance.gov/f/documents/102012_cfpb_fair-credit-reporting-act-fcra_procedures.pdf).

<sup>44</sup> 15 U.S.C. § 6502.

to operators of commercial websites and online services directed to children and also to general-audience services when the operator knows a child is providing personal information.<sup>45</sup>

COPPA regulates business practices surrounding the collection, use, and disclosure of children’s personal information. It requires parental consent, clear privacy notices, limits on data retention, and restrictions on sharing with third parties.<sup>46</sup> Its regulatory triggers attach to an operator’s handling of identifiable children’s data in providing the service. FTC guidance emphasizes that COPPA governs what information an operator collects, how it collects it, how it uses it, and how it discloses it—all tied to identifiable information associated with a child user.<sup>47</sup>

COPPA does not extend to downstream statistical or technical transformations once data no longer functions as identifiable personal information in the hands of a covered operator. A developer training a general-purpose LLM on broad web data, without operating a child-directed service or knowingly collecting children’s information from its own users, therefore falls outside the statute. In that setting, a model’s internal representations do not qualify as “personal information” under COPPA.

**Relevance to LLM Memorization:** COPPA is operator- and context-specific. Its duties attach to child-directed services and regulate the collection, use, retention, and disclosure of identifiable children’s data in providing that service. The statute fits front-end collection and deployment practices—consent flows, notices, retention limits, and onward disclosure—far better than it fits treating a general-purpose model’s parameters as children’s personal information. The most plausible COPPA issue therefore arises at the output stage: a covered operator could violate the statute by using an LLM to disclose a child’s identifiable information or by collecting such information through the interface without compliant consent and notice.

#### 4. *GLBA’s Financial-Institution Scope*

The Gramm–Leach–Bliley Act (GLBA) governs the privacy and security practices of financial institutions, broadly defined as entities “significantly engaged” in offering financial products or services to consumers.<sup>48</sup> FTC guidance makes the boundary explicit: GLBA applies to businesses such as lenders, check-cashing services, and financial advisers.<sup>49</sup>

---

<sup>45</sup> See Complying with COPPA: Frequently Asked Questions, FED. TRADE COMM’N, <https://www.ftc.gov/business-guidance/resources/complying-coppa-frequently-asked-questions> (last visited Feb. 16, 2026).

<sup>46</sup> 15 U.S.C. §§ 6502–03 (setting requirements for parental consent, notice, data minimization, and limits on use and disclosure of children’s information).

<sup>47</sup> See Children’s Online Privacy Protection Rule: A Six-Step Compliance Plan for Your Business, FED. TRADE COMM’N, <https://www.ftc.gov/business-guidance/resources/childrens-online-privacy-protection-rule-six-step-compliance-plan-your-business> (last visited Feb. 16, 2026).

<sup>48</sup> 16 C.F.R. § 313.3(k)(1); 15 U.S.C. § 6809(3)(A).

<sup>49</sup> 16 C.F.R. § 313.3(k)(1).

GLBA’s privacy framework centers on nonpublic personal information (NPI) arising from consumer financial relationships.<sup>50</sup> Financial institutions must provide privacy notices, limit disclosure of NPI to nonaffiliated third parties unless an exception applies, and implement administrative, technical, and physical safeguards to protect the information.<sup>51</sup> Implementing regulations focus on failures to safeguard NPI or improper disclosure of identifiable NPI, not on unrelated entities that merely process information touching on financial topics.<sup>52</sup>

Because GLBA obligations attach to financial institutions and to their handling of identifiable NPI, the statute does not extend to entities outside the financial-services context or to downstream statistical transformations of data occurring outside a regulated financial relationship. A general-purpose machine-learning model trained on web text does not become a financial institution simply because the corpus contains finance-related material. Likewise, statistical model parameters do not constitute “disclosure” or “sharing” of NPI under the statute.

**Relevance to LLM Memorization:** GLBA is a sectoral, entity-based regime focused on financial institutions’ treatment of customer financial information. Its core duties are operational: provide notices, limit certain disclosures of NPI, and maintain safeguards. The statute fits LLM deployment inside regulated institutions and their service-provider relationships, where the question is whether use of an LLM results in improper disclosure of identifiable NPI or inadequate safeguards. By contrast, GLBA does not plausibly treat a general-purpose model developer as a regulated financial institution or internal model parameters as NPI disclosure merely because training data included finance-related text. GLBA concerns arise when customer NPI enters, is processed by, or is output from an LLM in a manner that implicates safeguarding or disclosure obligations—not from background exposure during model training.

### 5. *VPPA’s Disclosure-Based Scope*

The Video Privacy Protection Act (VPPA) limits when a “video tape service provider” may disclose information identifying a consumer’s video-viewing history.<sup>53</sup> The statute defines such a provider as a business engaged in the rental, sale, or delivery of prerecorded video cassette tapes or similar audiovisual materials.<sup>54</sup> Courts have interpreted modern on-demand streaming services as falling within the statute’s scope.<sup>55</sup>

---

<sup>50</sup> See How to Comply with the Privacy of Consumer Financial Information Rule of the Gramm-Leach-Bliley Act, FED. TRADE COMM’N, <https://www.ftc.gov/business-guidance/resources/how-comply-privacy-consumer-financial-information-rule-gramm-leach-bliley-act> (last visited Feb. 16, 2026).

<sup>51</sup> *Id.*

<sup>52</sup> 16 C.F.R. pt. 314.

<sup>53</sup> 18 U.S.C. § 2710.

<sup>54</sup> 18 U.S.C. § 2710(a)(4).

<sup>55</sup> See *In re Hulu Priv. Litig.*, 86 F. Supp. 3d 1090, 1095 (N.D. Cal. 2015); *Buechler v. Gannett Co.*, No. CV 22-1464-CFC, 2023 WL 6389447, at \*2 (D. Del. Oct. 2, 2023).

Liability arises when a covered provider knowingly discloses PII linking an individual to specific video materials or services.<sup>56</sup> As with other federal privacy statutes, the VPPA turns on disclosure of identifiable information by a covered provider, not on internal technical processing within a general-purpose model.

**Relevance to LLM Memorization:** The VPPA is a narrow, disclosure-focused statute aimed at a specific context: video-service providers and the release of information connecting an identifiable person to particular viewing choices. Its trigger is the knowing disclosure of protected identifiers tied to video materials or services, not abstract data use. The statute therefore becomes relevant to LLM systems only when deployed within, or on behalf of, a covered streaming provider in a way that could reveal an individual's viewing history, e.g., through outputs, logs, or third-party sharing. By contrast, the VPPA does not plausibly treat a general-purpose model's internal parameters as protected PII or treat training exposure to audiovisual-related text as disclosure of viewing history.

## 6. *An Output- and Harm-Based Privacy Framework*

The FTC's "unfairness" authority is sometimes described as a more flexible federal privacy tool. Even so, liability under FTC doctrine requires a substantial consumer injury that consumers cannot reasonably avoid and that lacks countervailing benefits.<sup>57</sup> Internal model encodings do not meet that standard by themselves, because any injury depends on a downstream output. The mere existence of model weights capable of producing PII-like text therefore does not, standing alone, implicate the FTC's privacy or data-security framework. Even the federal government's broadest consumer-protection authority does not convert model parameters into regulated personal data.

Across the statutes surveyed, a consistent principle emerges. Liability arises when a system discloses identifiable information, is used in an eligibility decision, or otherwise enables unauthorized access or misuse in a regulated context.<sup>58</sup> The statutes attach duties to specific actors—health-care providers, consumer-reporting agencies, child-directed service operators, financial institutions, and video-service providers—and to defined activities such as collection, safeguarding, and disclosure. None treats a trained model's internal statistical structure as a legally cognizable repository of personal data.

The same logic applies to memorization. The presence of memorized PII within model weights is insufficient, by itself, to create liability. Legal risk materializes when outputs reveal identifiable information under circumstances that make the disclosure unlawful or when deployment practices violate statutory duties. Even state regimes that extend somewhat further, such as the California

---

<sup>56</sup> 18 U.S.C. § 2710(a)(3).

<sup>57</sup> 16 C.F.R. § 424.1.

<sup>58</sup> See 15 U.S.C. §§ 45(a), 45(n) (requiring substantial consumer injury for unfairness liability); 18 U.S.C. § 1030(a)(2) (imposing liability for unauthorized access to obtain information); 42 U.S.C. § 1320d-6(a); 45 C.F.R. § 164.502(a) (prohibiting unlawful use or disclosure of protected health information); 18 U.S.C. § 2701(a) (imposing liability for unauthorized access to stored communications). These statutes generally attach liability to unauthorized access, use, or disclosure of identifiable information, not to the mere existence of internal data representations.

Privacy Rights Act (CPRA, covered in Section IV.B), focus on business practices surrounding the handling of personal information, rather than the internal statistical architecture of a trained model.<sup>59</sup>

Taken together, U.S. privacy law reflects an output- and harm-focused framework. It evaluates how information is used, disclosed, or safeguarded in real-world interactions, not whether statistical parameters could theoretically encode personal data. Under that structure, treating LLM model weights as inherently regulated personal information is difficult to reconcile with existing doctrine.

## **B. The CPRA Regulates Data Practices and Outputs, Not Model Weights**

California’s privacy regime, governed by the California Privacy Rights Act (CPRA)<sup>60</sup> and administered by the California Privacy Protection Agency (CPPA),<sup>61</sup> differs from federal law in structure but not in its basic trigger. The statute establishes baseline consumer rights, creates a category of “sensitive personal information,” and authorizes a dedicated regulator. It does not impose a general lawful-basis requirement for data processing. Instead, it regulates business practices surrounding collection, use, retention, sharing, and security. Liability turns on proportionality, consumer choice, and risk, not on the mere existence of internal representations derived from personal data. The CPRA also excludes certain publicly available and lawfully obtained truthful information that is a matter of public concern, which affects how the statute treats widely disseminated information about public figures.<sup>62</sup>

This structure matters for memorization. The CPRA does not expressly treat machine-learning model weights as personal information, nor does it presume statistical parameters are regulated simply because personal data contributed to their formation. The statute defines personal information relationally—data that identifies, relates to, describes, or can reasonably be linked to a particular consumer or household. Its exclusion for publicly available and newsworthy information further narrows the set of regulated outputs. Internal parameters therefore fall outside the statute, absent reasonable linkability or output-stage disclosure.

Recent California developments nonetheless move beyond notice-and-opt-out compliance and toward system-oriented governance relevant to memorization risk. The CPPA emphasizes data minimization, limiting collection, use, retention, and sharing to what is reasonably necessary and

---

<sup>59</sup> See *How CPRA Defines Personal Information*, TRANSCEND (May 19, 2023), <https://transcend.io/blog/cpra-personal-information>; Cal. Civ. Code § 1798.140 (defining “personal information” as information that identifies, relates to, or could reasonably be linked to a consumer); *id.* §§ 1798.105, 1798.110 (requiring disclosure of categories of personal information collected and the purposes for which it is used, regulating business handling of personal information rather than internal technical representations not reasonably linkable to a consumer).

<sup>60</sup> Cal. Civ. Code § 1798.199.100 (2025).

<sup>61</sup> Cal. Civ. Code § 1798.199.10 (2025).

<sup>62</sup> Cal. Civ. Code § 1798.140(v)(1)(L).

proportionate to disclosed purposes.<sup>63</sup> Although framed in legal rather than technical terms, minimization addresses upstream conditions associated with memorization, including excessive retention, overbroad ingestion, and large-scale duplication of low-entropy personal data. In practice, the requirement encourages data curation and deduplication that can reduce verbatim or near-verbatim reproduction.

California has also adopted governance requirements that create compliance touchpoints for advanced AI systems, without relying on a database analogy. CCPA rules on cybersecurity audits and risk assessments require certain businesses to evaluate and document foreseeable privacy and security risks and to implement proportional safeguards.<sup>64</sup> These obligations are model-adjacent, rather than model-centric. They do not assume that an AI model stores personal data in the ordinary sense. Instead, they focus on whether system design, deployment, and safeguards reasonably address identifiable risks, including the possibility that sensitive information could appear in outputs under foreseeable use or misuse.

## V. Conclusion

This issue brief advances a narrower, empirically grounded understanding of memorization in large language models and urges greater care in mapping that concept onto privacy law. The technical literature consistently shows that memorization—verbatim or near-verbatim reproduction of training text—is rare relative to modern corpus scale, concentrated in low-entropy and highly duplicated material, and meaningfully reduced through established safeguards such as data curation, deduplication, and output filtering. Memorization is not a general feature of model behavior. It is a localized phenomenon that appears under specific conditions, often requiring targeted prompting.

U.S. privacy law largely tracks this reality. Across privacy torts and sector-specific statutes, liability attaches to disclosure, misuse, or failure to safeguard identifiable personal information in defined relationships and contexts. The law does not treat internal statistical representations as personal data merely because personal information may have contributed to their formation, and it does not equate probabilistic inference or hallucination with the exfiltration of stored records. A model's weights, standing alone, do not establish a privacy violation. What matters is whether identifiable information is actually disclosed, accessed, or used in a legally relevant way at the output or deployment stage.

California's framework is broader but points in a similar direction. The California Privacy Rights Act introduces data minimization, sensitive personal information, and risk-assessment obligations that expand oversight beyond federal sectoral statutes. Even so, California law remains focused on business practices, proportionality, and downstream sharing, rather than on treating models as

---

<sup>63</sup> See Applying Data Minimization to Consumer Requests, CAL. PRIV. PROT. AGENCY, ENF'T DIV., <https://coppa.ca.gov/pdf/enfadvisory202401.pdf> (last visited Feb. 15, 2026).

<sup>64</sup> See William E. Ridgway et al., *California Finalizes CCPA Regulations for Automated Decision-Making Technology, Risk Assessments and Cybersecurity Audits*, SKADDEN (Oct. 3, 2025), <https://www.skadden.com/insights/publications/2025/10/california-finalizes-cppa-regulations>.

repositories of personal data. In practice, the statute creates additional compliance obligations for AI systems without converting internal parameters into regulated records.

Several implications follow for policymakers. First, large language models are not databases. Their parameters encode probabilistic relationships rather than discrete entries, and memorization of personal information is atypical. Second, hallucination and inference raise different concerns from memorization and should not be treated as evidence of data leakage. Third, regulatory approaches that rely on database analogies or categorical assumptions about model behavior risk imposing disproportionate costs relative to the privacy harms at issue.

Accordingly, broad classification rules that treat all generated personal data—real or fabricated—as retrieved from storage risk overdeterrence. They encourage unnecessary restrictions on generative systems even when the risk of disclosing memorized personal information is low. A more coherent approach would align legal obligations with empirical realities by focusing on context, risk, and outputs. Such an approach better protects privacy, while preserving the substantial social value generative models can provide.