Yale Journal of Law & Technology
Volume 28, Symposium Issue

The Hypocrisy of Data Governance

Zubair Shafiq, Olivia Figueira, Athina Markopoulou, Woodrow Hartzog & Michael Lavine

"Data governance" is an empty term, like a Rorschach inkblot just waiting to be filled with meaning. Tech companies take advantage of this ambiguity to craft narratives about their data-governance capabilities to fit their audience and purpose. On one hand, tech companies brag about their data-governance capabilities when it fits their business model (for example, to advertisers) and public image (for example, to their customers). On the other hand, tech companies claim that meaningful data governance is challenging or impossible when accountability is demanded.

In this Article, we argue that tech companies systematically misrepresent or selectively ignore their data-governance capabilities. To demonstrate our point, we present two case studies showing how tech companies adopt inconsistent and self-serving positions when it comes to the treatment of consumers' personal information. First, we show examples where tech companies actively identify children to deliver personalized advertising and content recommendations but disclaim the knowledge or ability to identify children when legal obligations attach. Second, we show how tech companies commonly claim they do not know whether the information collected by their tracking tools is protected health information (PHI) under HIPAA, even though standard techniques enable such classification.

We conclude this article by arguing for a more sustained critique and skepticism of the concept and implementation of data governance. Lawmakers could better scrutinize what constitutes reasonable efforts under existing data protection rules, they could better tailor new rules to the data governance capabilities of tech companies, and finally, lawmakers could better scrutinize the use of the term "data governance" as an efficacy claim within the law of consumer protection.

Article Contents

Introduction

"Data governance" is a Rorschach inkblot test. That is, it is an intentionally vague and amorphous term whose meaning varies depending on the context and intended audience. Tech companies consistently use this vagueness to their advantage. They boast about their data-governance initiatives' capabilities when it suits their business interests but claim that they lack some of those same capabilities when regulators try holding them accountable.

As much as it is used across the tech industry, data governance is not a term of art. It has never been consistently defined in legislation, regulatory frameworks, or court opinions outside of vague gesturing towards information management. Even colloquially it has no universally accepted meaning. To public policy scholars, data governance typically refers to macro-level "norms, policies, and rules governing" digital information.[1] To the European Union and some U.S. state governments, data governance refers to mechanisms that facilitate data sharing between different organizations for public-interest purposes.[2] To tech companies, it means something different.[3]

Generally, this kind of ambiguity is not a problem. Terms like "big data" and "social media" are similarly amorphous and get refined in other ways in law and policy. However, tech companies have begun to exploit the ambiguity of "data governance" to avoid accountability for their information

---

[1] *FAQ*, DIGIT. TRADE & DATA GOVERNANCE HUB, https://datagovhub.elliott.gwu.edu/faq/ [https://perma.cc/2E9L-ZUBM] (last visited Jan. 27, 2024).

[2] *See, e.g.*, *European Data Governance Act*, EUR. COMM'N (Oct. 10, 2024) https://digital-strategy.ec.europa.eu/en/policies/data-governance-act [https://perma.cc/5JCS-MRME]; 8 COLO. CODE REGS. § 1501-7:4 (2024); HAW. CODE R. § 11-188-3 (2024).

[3] *See, e.g.*, *What is data governance?,* MICROSOFT, *https://azure.microsoft.com/en-us/resources/cloud-computing-dictionary/what-is-a-data-governance* [https://perma.cc/B9Z2-HW87] ("The definition of data governance includes the collection of processes, policies, roles, metrics, and standards that ensures an effective and efficient use of information.").

practices. To make matters worse, tech companies also take advantage of ambiguities not only with respect to "data governance" as a term, but also with their data-governance capabilities. Tech companies often simultaneously portray their data-governance programs as highly capable in the context of value maximization *and* incapable in the context of privacy compliance. This hypocrisy is the problem this Article seeks to examine.

For starters, organizations across the tech industry inconsistently define data governance. Some organizations use the term to refer to the rules and policies that cover the management of their data throughout its entire lifecycle, "from intake to disposal."[4] Others specifically distinguish data governance as a subdiscipline of data management.[5] In this view, data governance focuses only on managing already-collected data.[6]

Tech companies are also inconsistent about the relationship between data governance and other concepts like privacy, data security, and compliance. Google describes data security and access management as "core data governance competencies" and incorporates data privacy into its governance framework.[7] For many organizations, compliance needs are the main reason for adopting data-governance initiatives (and purchasing data-governance tools).[8] Indeed, data governance is often used as a euphemism for compliance efforts. Yet, in other circumstances, data governance refers to an organization's ability to reliably

---

[4] *What is data governance?,* MICROSOFT SEC., https://www.microsoft.com/en-us/security/business/security-101/what-is-data-governance-for-enterprise [https://perma.cc/9NJX-KU9V] (last visited Jan. 27, 2025); *see also Data Governance*, DATABRICKS, https://www.databricks.com/discover/data-governance [https://perma.cc/LSR6-QHA2] (last visited Apr. 29, 2025).

[5] Jim Holdsworth & Matthew Kosinki, *What is data governance?*, IBM, (Sep. 20, 2024), https://www.ibm.com/think/topics/data-governance [https://perma.cc/TT25-BARC].

[6] *Id.*

[7] *Principles and best practices for data governance in the cloud*, GOOGLE CLOUD, https://services.google.com/fh/files/misc/principles_best_practices_for_data-governance.pdf [https://perma.cc/X7ZT-Y39P] (last visited Apr. 29, 2025).

[8] RUPA MAHANTI, DATA GOVERNANCE AND COMPLIANCE: EVOLVING TO OUR CURRENT HIGH STAKES ENVIRONMENT 150 (2021).

leverage its data to achieve its strategic goals.[9] In other words, data governance is really about extracting value from data.

Microsoft, for instance, suggests that privacy and security are embedded in and are core goals of data governance.[10] But when marketing its enterprise data-management platforms, Fabric and Purview, it boasts three different suites of features for the three separate disciplines of "data governance," "data security," and "risk and compliance"—reflecting a vision of "governance" as a different function than security or compliance.[11] In its training materials for Purview, Microsoft explains data governance as a separate discipline than data security or compliance; one that is focused squarely on ensuring the accessibility and quality of an organization's data.[12]

In some ways, data governance's compliance and value-maximization functions are compatible. When organizations follow the fair information practices to process data, that data tends to be more accurate, benefitting both the data subject and those who seek to exploit the data. More than that, the data-governance tools used to extract data's value are often the very same tools that could enable robust compliance. One data-governance specialist suggests as much when he explained that:

> Data governance has a crucial role in compliance . . .[it] provides confidence for the business to act without fear of inadvertently

---

[9] Kimberly A. Houser & John W. Bagby, *Next-Generation Data Governance*, 21 DUKE L. & TECH. REV. 61, 69 (2024) ("Data governance is, first and foremost, a data management function to ensure the quality, integrity, security, and usability of the data collected by an organization. To corporations, data governance is a set of procedures and policies designed to both manage and monetize data.") (quotation omitted).

[10] *What is data governance?, supra* note 3.

[11] *Microsoft Purview*, MICROSOFT, https://learn.microsoft.com/en-us/purview/ [https://perma.cc/FP4Q-6M7R] (last visited Apr. 29, 2025).

[12] *Data governance with Microsoft Purview*, MICROSOFT, https://learn.microsoft.com/en-us/purview/data-governance-overview [https://perma.cc/VQ7U-ZG4X] (last visited Apr. 29, 2025) ("Data governance is how you ensure that the data you use in your business operations, reports, and analysis is discoverable, accurate, trusted, and protected. With ever-evolving regulations . . . data governance is as critical as data security.").

> breaking the rules or the law. Similarly, the same approaches, techniques, and technology around data governance for compliance can be applied for profit as they shine a light on the business operations.[13]

Yet tech companies tend to describe the capabilities of these "approaches, techniques, and technolog[ies]" differently depending on their audience.[14] When promoting data-governance services and products to other businesses, data governance facilitates effective compliance and security solutions. When working with advertisers, data governance allows for targeted, fine-grained identification of different classes of users and information. When promoting consumer products, companies claim that data governance provides a high degree of protection for vulnerable users and sensitive data. But when regulators attempt to hold tech companies accountable for alleged privacy violations, those same data-governance tools are suddenly inadequate to allow compliance with the law.

Tech companies tout their data governance when it stands to benefit these companies' bottom lines, by selling their products, attracting advertisers, or cultivating social capital through "privacy-washing." Recent advances in AI are only accelerating the hype around companies' ability to analyze and utilize data. By contrast, tech companies tend not to disturb assumptions about their capabilities when it increases liability risk or compliance costs. Nowhere is this hypocrisy more obvious than in the context of sensitive data.

Despite the notable lack of a federal comprehensive privacy law, Congress has managed to pass legislation introducing protections for narrower categories of especially vulnerable information and information subjects. In this Article, we focus on the following two examples. First, federal law imposes more stringent requirements on the collection and processing of information from young children online.[15] Second, it imposes more stringent requirements on the

---

[13] MAHANTI, *supra* note 8, at 112.

[14] *See id.*

[15] *See* Children's Online Privacy Protection Act of 1998, codified at 15 U.S.C. § 6501 *et seq.* (2024).

collection and processing of certain information related to individuals' healthcare.[16] Young children's data and health-related data warrant increased protection, in part, because they can leave data subjects in a particularly vulnerable position. However, these are particularly valuable forms of data for tech companies, which are incentivized to both maximize the value they can extract from this data and, where possible, avoid triggering the heightened legal protections these types of data receive. They do this by leveraging the ambiguity of data governance and their data-governance capabilities. Tech companies are able to identify these types of sensitive data when selling it to advertisers or optimizing their own products and services, yet feign ignorance when knowledge of that data might trigger costly obligations.

This Article proceeds by examining how the hypocrisy of data governance plays out in the context of children's data and health-related data, respectively. It then concludes by arguing for law and policy responses that tackle these ambiguities head on and demand consistency from tech companies.

I. The Case of Identifying Children

Children are increasingly using mobile devices and social media platforms more intensively and at younger ages.[17] Companies have a strong monetary incentive to allow, engage, and target children on their platforms. For example, expenditures on online advertising to children have increased dramatically in recent years and are expected to have exceeded $6 billion in 2025.[18] In 2021, the U.S. made up the largest share

---

[16] *See* Health Insurance Portability and Accountability Act of 1996, codified at 42 U.S.C. § 1320d *et seq.* (2024).

[17] THE COMMON SENSE CONSENSUS: MEDIA USE BY KIDS ZERO TO EIGHT, COMMON SENSE (2025), https://www.commonsensemedia.org/sites/default/files/research/report/2025-common-sense-census-web-2.pdf [https://perma.cc/JX3S-BCNB].

[18] *Spending on digital advertising to children worldwide from 2021 to 2031*, STATISTA (July 3, 2023) https://www.statista.com/statistics/1326893/children-digital-advertising-spending-worldwide/ [https://perma.cc/SBS2-8QEF].

of worldwide spending on online advertising to children, totaling $1.08 billion, or 37% of such advertising.[19]

The Children's Online Privacy Protection Act of 1998 (COPPA) is the federal law for protecting the online privacy of children under thirteen years of age.[20] A year after its passage, the Federal Trade Commission (FTC) issued the "COPPA Rule" to implement COPPA's privacy requirements.[21] The COPPA Rule regulates the collection, use, and disclosure of personal information of children online.[22] Specifically, it requires online-service operators to obtain verifiable parental consent before they collect, use, or disclose children's personal information if either (1) their service is directed to children or (2) the operator has "actual knowledge" that they are collecting personal information from a child.[23]

Because of two conflicting objectives (that is, business objectives and compliance with privacy laws), companies often make conflicting statements about their capabilities and engage in poor privacy practices, including tracking, targeted advertising, and collection and sharing of children's personal information without parental consent. Companies often claim to lack "actual knowledge" of users' ages or do not consider

---

[19] *Spending on digital advertising to children worldwide in 2021, by region*, STATISTA (July 3, 2023) https://www.statista.com/statistics/1326908/children-digital-advertising-spending-region-worldwide/ [https://perma.cc/D2EM-BV3L].

[20] Children's Online Privacy Protection Act of 1998, codified at 15 U.S.C. § 6501 *et seq.* (2024).

[21] Press Release, FED. TRADE COMM'N, New Rule Will Protect Privacy of Children Online (Oct. 20, 1999), https://www.ftc.gov/news-events/news/press-releases/1999/10/new-rule-will-protect-privacy-children-online [https://perma.cc/T4LG-LHMZ].

[22] "Personal information" is defined in COPPA as "individually identifiable information about an individual collected online," including name, address, online contact information, screen or user name, telephone number, social security number, persistent identifiers, photograph, video, or audio of a child's image or voice, geolocation, and information about the child or their parents that can be combined with another identifier. *See* 16 C.F.R. § 312.2 (2025).

[23] *See* 16 C.F.R. § 312.3 (2025). However, Section 312.5(c) provides a few exceptions to the prior parental consent requirement including, *inter alia*, processing "to protect the safety of a child" or collecting a persistent identifier required for "internal operations" like maintaining functionality, security, or contextual advertising. 16 C.F.R. § 312.5(c) (2025).

their platforms to be "child-directed" and thus do not need to comply with COPPA. At the same time, some companies are often actively identifying or targeting content and advertising to children for monetization purposes.[24]

This case study demonstrates data governance's double standards, i.e., the fact that statements by companies about their capabilities and practices vary depending on the audience and business interests. To their customers, companies market their ability to protect children's privacy and safety. To regulators, companies claim that they simply cannot know when a child is using their services. However, age identification is technically possible today and is widely used to target them with personalized content or advertising. Indeed, when talking to internal product groups, customers, and advertisers, companies discuss and offer their capabilities to precisely target children for monetization purposes through increased engagement and targeted advertising.

What companies say to the public *(The message is along the lines that they protect children on their platforms.)*. Companies often provide privacy features and protections for children.

For example, Google Ads prohibits targeted advertising to children on Google's platforms.[25] Google also provides a service called "Family Link", which is marketed as a way to "[h]elp keep your family safer online" when using Google products like Android and ChromeOS devices.[26] Family Link includes screen time limits, permissions management, such as for websites and extensions on children's devices, and location sharing. Google also provides security and privacy features of their "Google Workspace for Education" tools and Chromebook devices for educational settings. For example,

---

[24] Stephen Morris & Hannah Murphy, *Google and Meta struck secret ads deal to target teenagers*, FIN. TIMES (Aug. 8, 2024), https://www.ft.com/content/b3bb80f4-4e01-4ce6-8358-f4f8638790f8 [https://perma.cc/3KX9-J9SX].

[25] *About demographic targeting*, GOOGLE ADS HELP, https://support.google.com/google-ads/answer/2580383?hl=en&sjid=3257305426111513472-NC [https://perma.cc/844N-DJ8D] (last visited Apr. 28, 2025).

[26] *Family Link*, GOOGLE, https://families.google/familylink/ [https://perma.cc/5LA9-MZ3N] (last visited Apr. 28, 2025).

Google claims it keeps its education products "private and secure" by operating their "own servers and services," prohibiting targeted advertising, maintaining compliance with privacy laws like COPPA, being transparent about the collection of users' data, providing "age-based content" and "safer browsing settings" on their apps and devices,[27] and setting default privacy settings for educational devices used by young children in schools.[28] As another example, Meta has developed "child protection tools" for their platforms, Instagram and Facebook, such as tools to identify and remove child sexual abuse material and to prevent child sexual harassment by limiting communication between teen accounts and potentially unsafe accounts.[29] Furthermore, TikTok provides a separate, "age-appropriate" version of their service specifically for children under thirteen, as described next, in an effort to better protect children (e.g., user accounts are private and users cannot access comments on videos nor message other users).[30]

What companies say to law enforcement *(The message is along the lines that, to the best of their knowledge, there are no children on the regular platforms.).* Some services provide separate versions of their platforms specifically for children under thirteen that are meant to provide more privacy protections under COPPA, such as YouTube Kids by YouTube[31] and TikTok's "Kids Mode" (officially called "Under 13 Experience").[32]

---

[27] *Privacy & Security FAQ*, GOOGLE FOR EDUC., https://edu.google.com/intl/ALL_us/our-values/privacy-security/frequently-asked-questions/ [https://perma.cc/6BJQ-V8YD] (last visited Apr. 28, 2025).

[28] *Chromebooks Overview*, GOOGLE FOR EDUC., https://edu.google.com/intl/ALL_us/chromebooks/overview/ [https://perma.cc/48H5-4XMP] (last visited Apr. 28, 2025).

[29] *Online Child Protection,* META SAFETY CTR., https://www.meta.com/safety/topics/online-child-protection/ [https://perma.cc/C77E-GYWF] (last visited Sep. 27, 2025).

[30] *Tiktok Under 13 Experience*, TIKTOK, https://support.tiktok.com/en/safety-hc/account-and-user-safety/tiktok-under-13-experience [https://perma.cc/4RYT-WHGC] (last visited Sep. 27, 2025).

[31] YOUTUBE KIDS, https://www.youtubekids.com/ [https://perma.cc/DP46-EVQB] (last visited Apr. 28, 2025).

[32] *Tiktok Under 13 Experience*, *supra* note 30.

Others, like Instagram[33] and Facebook,[34] simply do not allow children under thirteen to use their services. Historically, Instagram and Facebook relied on age screens that ask the user to input their birth date during account creation. If a user represented themselves as younger than thirteen, Facebook and Instagram would prevent them from creating an account. Tech companies have repeatedly used this age screening process to claim their services are not child-directed and to deny "actual knowledge" of children under thirteen using their services.

Either of the above approaches allows companies to treat users on their regular platforms as adults. This, of course, relies on the user's honesty when providing their age. But children can, and often do, lie during the account creation process.[35] Almost 40% of children between the ages of eight and twelve use social media.[36] Yet, because tech companies require users to input their age, they can skirt COPPA requirements by claiming they lack "actual knowledge" of children on their platforms.

As outlined later in this section, investigative journalism and comprehensive audits conducted by researchers have provided strong evidence that companies may be taking advantage of this potential loophole to skirt COPPA requirements, even after significant FTC settlements on these issues, e.g., as in the cases of YouTube and TikTok.[37] When it

---

[33] *Terms of Use*, INSTAGRAM, https://help.instagram.com/termsofuse [https://perma.cc/BB5U-DMN2] (last visited Apr. 28, 2025).

[34] *Terms*, FACEBOOK, https://www.facebook.com/terms [https://perma.cc/8W99-EJW4] (last visited Apr. 28, 2025).

[35] danah boyd, Eszter Hargittai, Jason Schultz & John Palfrey, *Why Parents Help Their Children Lie to Facebook About Age: Unintended Consequences of the 'Children's Online Privacy Protection Act'*, 16 FIRST MONDAY (Nov. 2011).

[36] Carol Vidal & Jennifer Katzenstein, *Social Media and Mental Health in Children and Teens*, JOHN HOPKINS MED., https://www.hopkinsmedicine.org/health/wellness-and-prevention/social-media-and-mental-health-in-children-and-teens [https://perma.cc/CJ9G-QDM2] (last visited Apr. 28, 2025).

[37] Press Release, FED. TRADE COMM'N, *Google and YouTube Will Pay Record $170 Million for Alleged Violations of Children's Privacy Law* (Sept. 4, 2019), https://www.ftc.gov/news-events/news/press-releases/2019/09/google-youtube-will-pay-record-170-million-alleged-violations-childrens-privacy-law [https://perma.cc/79YN-VB52]

comes to compliance and protecting children's privacy, they claim to not have "actual knowledge" of children on their platforms or do not fall under the definition of "child-directed." However, tech companies are capable of identifying children for targeting personalized content (for engagement) and advertising,[38] i.e., for monetization purposes, as described next.

What companies say and do internally (*The message is along the lines that they need to engage children—the next generation of users*.). Companies may say they protect children, but it is important to look at their actual practices. While platforms like Facebook and Instagram are restricted to users ages thirteen and older, internal Facebook documents demonstrate the company's interest in increasing their products' appeal among children under thirteen.[39] In 2021, whistleblower Frances Haugen provided journalists with internal Facebook documents that specifically highlight the company's interest in introducing features for children under thirteen, organized by "user maturity" or age cohort.[40] For example, Facebook was interested in introducing features for children ages zero to four, ages five to nine, and tweens ages ten to twelve. One 2018 document stated, "[w]ith the ubiquity of tablets and phones, kids are getting on the internet as young as six years old. We can't ignore this and we have a responsibility to figure it out."[41] In 2017, Facebook introduced

---

[hereinafter Google & YouTube COPPA Violation Press Release]; Press Release, FTC, FTC Investigation Leads to Lawsuit Against TikTok and ByteDance for Flagrantly Violating Children's Privacy Law (Aug. 2, 2024) https://www.ftc.gov/news-events/news/press-releases/2024/08/ftc-investigation-leads-lawsuit-against-tiktok-bytedance-flagrantly-violating-childrens-privacy-law [https://perma.cc/Z8UY-MPKT] [hereinafter TikTok & ByteDance COPPA Lawsuit Press Release].

[38] Google & YouTube COPPA Violation Press Release, *supra* note 37; Morris et al., *supra* note 24.

[39] Georgia Wells & Jeff Horowitz, *Facebook's Effort to Attract Preteens Goes Beyond Instagram Kids, Documents Show*, WALL STREET J. (Sep. 28, 2021), https://www.wsj.com/articles/facebook-instagram-kids-tweens-attract-11632849667 [https://perma.cc/GX5A-HY6Y].

[40] Ryan Mac & Cecilia Kang, *Whistle-Blower Says Facebook 'Chooses Profits Over Safety'*, N.Y. TIMES (Oct. 3, 2021), https://www.nytimes.com/2021/10/03/technology/whistle-blower-facebook-frances-haugen.html [https://perma.cc/94ZQ-5BXU].

[41] *Id.*

Messenger Kids as a child-specific platform for users ages six to twelve, but Facebook researchers found that interest in the service waned for users ages ten and older, and that these tween users were not yet interested in Instagram. In a written statement responding to the *Wall Street Journal*'s reporting on these internal documents, head Instagram executive Adam Mosseri explained that "[l]ike all technology companies, of course, we want to appeal to the next generation, but that's entirely different from the false assertion that we knowingly attempt to recruit people who aren't old enough to use our apps."[42]

What companies say to advertisers (*Essentially, promising that they can accurately target ads to children or adolescents.*). Instagram's and Facebook's responses to these internal documents directly contradict the findings of a more recent investigation surrounding advertising deals between Google and Meta. In 2024, the *Financial Times* found through internal documents from Google and Meta that Google secretly allowed Instagram to target YouTube advertisements to users on their platform labeled as "unknown." Google allegedly knew that these "unknown" users were primarily under eighteen years old.[43] Google's terms prohibit targeted advertising to users under eighteen years old.[44] Thus, the Meta and Google deal appears to have bypassed Google's own policy to serve the companies' business interests. Google's advertising website states, "Google Ads can't know or infer the demographics of all people. 'Unknown' refers to people whose age, gender, parental status, or household income we haven't identified."[45]

The *Financial Times* argues that Google knew that this category contained children because they had separated out all other age groups eighteen and up. When coupled with users' online behaviors, such as their downloaded apps and browsing activity, they could use this information to confidently determine that the "unknown" category contained children.[46] Google did not explicitly deny these claims, stating instead that

---

[42] *Id.*
[43] Morris et al., *supra* note 24.
[44] *Id.*
[45] *About demographic targeting*, *supra* note 25.
[46] Morris et al., *supra* note 24.

"we'll also be taking additional action to reinforce with sales representatives that they must not help advertisers or agencies run campaigns attempting to work around our policies."[47] Google emphasized that their "safeguards" protecting children worked properly because no users self-identified as under-18 were directly targeted.[48]

FTC settlements regarding COPPA violations. In 2019 and 2024, YouTube and TikTok, respectively, reached settlements with the FTC for allegedly violating COPPA by collecting personal information from children without parental consent.[49]

In YouTube's case, they allegedly marketed their platform's popularity among children under thirteen to potential clients and advertisers, as shown by their popular child-directed channels.[50] Despite knowing that these channels were child-directed, and in some cases were also featured in their separate YouTube Kids app, the complaint from the FTC and New York Attorney General alleged that YouTube served targeted advertisements and collected personal information from channel viewers without parental consent.[51] As part of the settlement, YouTube was required to improve the transparency and labeling of child-directed content on their platform by enabling channel owners to identify whether their content is child-directed and ensure COPPA compliance.[52] YouTube also stated in an announcement following the settlement that they would "treat data from anyone watching children's content on YouTube as coming from a child, regardless of the age of the user."[53] However, despite the FTC settlement, researchers at Adalytics found that as of July 2023, YouTube appeared to be tracking viewers of "made for kids"

---

[47] *Id.*
[48] *Id.*
[49] Google & YouTube COPPA Violation Press Release, *supra* note 37.
[50] Complaint, New York v. Google, LLC, No. 1:19-cv-2642 (D.D.C Sep. 4, 2019).
[51] *Id.*
[52] Google & YouTube COPPA Violation Press Release, *supra* note 37.
[53] Susan Wojcicki, *An update on kids and data protection on YouTube*, YOUTUBE (Sep. 4, 2019), https://blog.youtube/news-and-events/an-update-on-kids/ [https://perma.cc/ZR3B-NENH].

videos through cookies and persistent identifiers and serving targeted advertising on "made for kids" channels.[54]

In 2024, the FTC alleged that TikTok violated COPPA by "knowingly allowing millions of children under 13 on their platform designated for users 13 and older" and had continued to collect personal information from such users without parental consent.[55] TikTok allegedly allowed children under thirteen to create accounts and access the platform without any age attestation or parental consent, such as through third-party sign-on services, and such accounts were labeled as "age unknown." The complaint states that TikTok classified millions of such accounts as "age unknown."[56] Furthermore, the complaint alleges that TikTok's "Kids Mode" service, which is meant for children under thirteen, violated COPPA by collecting and using personal information beyond what was necessary and without parental consent.[57]

Company practices revealed by research studies. In addition to reports and FTC settlements regarding how companies handle users under thirteen and their personal information, researchers have also gathered evidence through comprehensive audits of child-directed services and content. Out of nearly 6,000 Google Play Store applications, a majority of applications were potentially noncompliant with COPPA due to their usage of third-party software development kits (SDKs), with 19% collecting personally identifiable information through SDKs that are prohibited in child-directed applications and 66% transmitting persistent and non-resettable identifiers.[58] Among 2,000 of the most popular child-directed websites, which were identified through a machine-learning classifier among thousands of popular websites, 90% embedded at least one tracker in their website and 27% contained targeted advertisements, some of which contained

---

[54] *Are YouTube Advertisers Inadvertently Harvesting Data From Millions of Children?*, ADALYTICS, https://adalytics.io/blog/are-youtube-ads-coppa-compliant [https://perma.cc/7MJ9-GSCF] (last visited Apr. 28, 2025).
[55] TikTok & ByteDance COPPA Lawsuit Press Release, *supra* note 37.
[56] *Id.*
[57] *Id.*
[58] Irwin Reyes et al., *"Won't Somebody Think of the Children?": Examining COPPA Compliance at Scale*, 3. PROC. ON PRIV. ENHANCING TECHS. 63 (2018).

inappropriate or sexually-explicit content.[59] Researchers have found that popular general-audience services that are directed to users of all ages, including Duolingo, Minecraft, Quizlet, Roblox, TikTok (including TikTok's "Kids Mode"), and YouTube (including YouTube Kids), did not significantly alter their data-collection and sharing practices between child, adolescent, and adult users, even when the platforms had "actual knowledge" of users' ages through account creation procedures.[60] Additionally, many of the platforms began collecting and sharing personal information about users before user age disclosure and parental consent, demonstrating a potential loophole surrounding the "actual knowledge" requirement of user age in COPPA.[61]

In addition, Figueira et al. further investigated TikTok's "Kids Mode" (TKM) and found that TKM's content is predominantly not child-directed, based on COPPA's definition of child-directed content, and is highly repetitive, which indicates a surprisingly limited TKM content inventory. Furthermore, TKM contains some sexually explicit and profane content, which is inappropriate for children.[62] In contrast, there exists an abundance of child-directed content on TikTok's regular mode that could be curated for TKM.[63] These findings cast doubt on the purpose of TKM as a service, as it seems to be "for kids" only by name. If TKM is not engaging for children due to its lack of child-directed content and overall content inventory, children may abandon TKM for riskier platforms, such as TikTok's regular mode, where they would have to lie about their age to gain access and be exposed to the aforementioned privacy risks.

---

[59] Zahra Moti et al., *Targeted and Troublesome: Tracking and Advertising on Children's Websites*, 2024 IEEE SYMPOSIUM ON SEC. & PRIV. 1517.

[60] Olivia Figueira et al., *DiffAudit: Auditing Privacy Practices of Online Services for Children and Adolescents*, IMC '24: PROC. 2024 ACM ON INTERNET MEASUREMENT CONF. 488 (2024).

[61] *Id.*

[62] Olivia Figueira et al., When Kids Mode Isn't For Kids: Investigating TikTok's "Under 13 Experience" (June 30, 2025) (unpublished manuscript). https://arxiv.org/abs/2507.00299 [https://perma.cc/HD2H-VCXP].

[63] Martin Hilbert et al., *#BigTech @Minors: Social Media Algorithms Have Actionable Knowledge About Child Users and At-Risk Teens*, 103 TELEMATICS & INFORMATICS 102341 (2025).

Other researchers have found that targeted advertising to children under thirteen is still possible on YouTube and that advertisers continue to employ such advertisements through placement-based targeting[64] on child-directed videos on YouTube.[65] While Google states that they do not allow advertising to users under eighteen, the study found that placement-based advertising on YouTube allows advertisers to select the videos or channels on which to place their ads, and thus they may select child-directed videos and channels.

Part of the solution: age gating is technically possible. For online services that do not allow children on their platforms, companies claim to check for underage accounts and remove them from their platforms,[66] which implies the technical capability to do so. Some companies also enable users to report other users that they think may be under thirteen years of age, such as on TikTok,[67] Instagram,[68] and Facebook.[69] Users can appeal these decisions and verify their age using more sensitive information. For example, TikTok users have several options to verify that they are thirteen and older, including submitting a photo of themselves with government-issued identification, credit-card authorization, facial scanning, or, if the user is between thirteen to seventeen years old, a parent or guardian can provide credit-card information or submit a photo of them

---

[64] *Placement-based targeting*, GOOGLE, https://support.google.com/google-ads/answer/99502 [https://perma.cc/FQ7B-WE6U] (last visited Jan. 29, 2026).

[65] Tinhinane Medjkoune et al., *Marketing to children through online targeted advertising: Targeting mechanisms and legal aspects*, PROC. 2023 ACM SIGSAC CONF. ON COMP. AND COMMC'NS SEC. 180 (2023).

[66] *Age Requirements on Google Accounts*, GOOGLE, https://support.google.com/accounts/answer/1350409 [https://perma.cc/M556-5HXG] (last visited Apr. 28, 2025); *Underage Appeals on TikTok*, TIKTOK, https://support.tiktok.com/en/safety-hc/account-and-user-safety/underage-appeals-on-tiktok [https://perma.cc/4C83-BTWC] (last visited Apr. 28, 2025); *Report a Child Under 13 on Instagram*, FACEBOOK, https://www.facebook.com/help/instagram/517920941588885/?helpref=uf_share [https://perma.cc/2BWS-GKXT] (last visited Apr. 28, 2025).

[67] *Underage Appeals on TikTok*, *supra* note 66.

[68] *Report a Child Under 13 on Instagram*, *supra* note 66.

[69] *Report the Account of a Person who Doesn't Meet Our Age Requirements*, FACEBOOK, https://www.facebook.com/help/contact/209046679279097 [https://perma.cc/5Y98-YZA2] (last visited Apr. 28, 2025).

with the user.[70] On the other hand, if a child under thirteen wishes to create a Google account, Google requires a parent's authorization and credit-card verification to do so.[71] Additionally, other third-party companies, such as Yoti, Incode, and VerifyMyAge, have been developing tools for age estimation based on facial scanning, but such approaches are largely viewed as an invasion of privacy because they require sensitive information[72] and they can also make mistakes. In February 2025, Google announced that they would be developing a machine-learning-based age-estimation approach for their platforms to determine whether a user was older or younger than eighteen years old, and subsequently provide more "age-appropriate experiences."[73] They did not mention whether they would attempt to identify more specific age ranges, such as children under thirteen, nor did they mention how they would use this approach to protect children.

In 2022, Meta announced their artificial intelligence-based model to predict users' ages, specifically "whether someone is an adult (18 and over) or a teen (13–17)."[74] Meta did not mention whether they would use this to estimate the ages of users under thirteen, but they disclosed that this information would be used to prevent adolescents from accessing adult

---

[70] *Underage Appeals on TikTok*, *supra* note 66.

[71] *Age Requirements on Google Accounts*, *supra* note 66.

[72] Drew Harwell, *A booming industry of AI age scanners, aimed at children's faces*, Washington Post (Aug. 7, 2024), https://www.washingtonpost.com/technology/2024/08/07/face-scanning-kids-online-privacy/ [https://perma.cc/QH4H-CFDA].

[73] Jen Fitzpatrick, *New digital protections for kids, teens and parents*, GOOGLE (Feb. 12, 2025), https://blog.google/technology/families/google-new-built-in-protections-kids-teens/ [https://perma.cc/55HM-DBHA]. Note that Google also stated, "As we continue to find new ways to deliver age-appropriate safeguards, one of the most complex challenges is understanding the age of the user. This year we'll begin testing a machine learning-based age estimation model in the U.S. This model helps us estimate whether a user is over or under 18 so that we can apply protections to help provide more age-appropriate experiences. We'll bring this technology to more countries over time." *Id.*

[74] Erica Finkle et al., *How Meta Uses AI to Better Understand People's Ages on Our Platforms*, META (June 22, 2022), https://tech.facebook.com/artificial-intelligence/2022/6/adult-classifier/ [https://perma.cc/MWY8-NYFL].

features, such as Facebook Dating.[75] Users would be able to verify their age in other ways if the model is incorrect, such as through government-issued identification or a facial scan, and their model is trained on profile information, including when the account was created and behaviors on Meta platforms, such as interaction with other users and content.[76] In 2024, Meta also released a report regarding their efforts to protect teens on Instagram, sharing that they would begin testing their age-estimation system in 2025, specifically to identify users over or under eighteen years old.[77]

In summary, age estimation based on online activity is technically possible and continues to be developed using the data companies already have access to and regularly collect on their platforms. However, whether these age-estimation systems will be actually used to protect children online, such as by removing underage accounts, is yet to be seen.

Interestingly, these techniques seem to be already in use for various purposes, and alternative age-assurance methods continue to be developed. A group of researchers has found that social-media recommendation algorithms, including those within YouTube, TikTok, and Instagram, are capable of personalizing content based on age through behavioral indicators, and thus may possess "actionable knowledge" of users' ages.[78] Others have explored this issue in the context of age-estimation and verification mechanisms that can enable services to move away from simple age attestation, which users can easily lie about.[79] Age-estimation techniques involve online services inferring users' ages, such as through biometric data (for example, facial features, voice, and gait), capacity and knowledge testing, and profiling based on attributes that can

---

[75] *Id.*
[76] *Id.*
[77] *How Research and Consultation Informed Instagram Teen Accounts*, META TRUST, TRANSPARENCY & CONTROL LABS (Mar. 26, 2025), https://www.ttclabs.net/report/how-research-and-consultation-informed-instagram-teen-accounts-a-new-protected-experience-for-teens-guided-by-parents [https://perma.cc/U6RA-773A].
[78] Hilbert et al., *supra* note 63.
[79] Noah Apthorpe, Brett M. Frischmann & Yan Shvartzshnaider, *Online Age Gating: An Interdisciplinary Evaluation* (Aug. 1, 2024) (unpublished manuscript), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4937328 [https://perma.cc/TH9S-RYMC].

be learned through interacting with an application (including time spent on certain pages, interests, location, and friends). While such techniques may seem to increase data collection about users and enable continued user tracking, prior research has demonstrated that online services are already engaging in such behaviors and are capable of personalizing content by age.[80] Furthermore, age verification, which seeks to confirm the age of a user through a trusted third party, is common, such as through government-issued identification or a credit card. However, these approaches pose further risks to user privacy, and thus researchers continue to study more sophisticated cryptography-based, privacy-preserving approaches, such as anonymous credential authentication systems that incorporate public-key cryptography, blind digital signatures, and zero-knowledge proofs.[81]

In summary, complying with COPPA should not be technically difficult today. First, determining whether a service is child-directed is not difficult. The FTC has issued guidelines regarding how to comply with COPPA. This guidance includes a list of factors used to determine whether an online service is child-directed, such as the use of animated characters, music, age of models, and advertising directed to children.[82] These compliance guidelines also include detailed explanations of responsibilities of general audience platforms, mixed-audience platforms, involvement of third parties, and much more.[83] Additionally, researchers have developed machine-learning classifiers to automatically categorize services as child-directed or not and thus whether it must comply with COPPA.[84] Second, as discussed prior, it is possible through various techniques to determine users' ages,[85] and companies are already doing it internally.[86] Thus, companies should be easily able to

[80] Hilbert et al., *supra* note 63.
[81] Apthorpe et al., *supra* note 79.
[82] Item D.1 of the FTC's list of frequently asked COPPA questions elaborates on this. *Complying with COPPA: Frequently Asked Questions*, FTC https://www.ftc.gov/business-guidance/resources/complying-coppa-frequently-asked-questions [https://perma.cc/Q3VM-3D2Z] (last visited Apr. 28, 2025).
[83] *Id.*
[84] Moti et al., *supra* note 59.
[85] Apthorpe et al., *supra* note 79.
[86] Hilbert et al., *supra* note 63; Morris et al., *supra* note 24.

determine whether they must comply with COPPA through these methods and by consulting with the detailed FTC guidelines on complying with COPPA.[87]

Part of the solution: proposed legislation. The industry also continues to push back against proposed children's privacy and safety legislation at both the state and federal levels. The California Age-Appropriate Design Code Act (CAADCA), modeled after the United Kingdom's enacted Age Appropriate Design Code (AADC),[88] aims to protect users under 18 years old and would apply to any entity that is "likely to be accessed by a minor," contrasting with COPPA's "actual knowledge" requirement.[89] CAADCA would prohibit companies from collecting, selling, sharing, and retaining personal information about minors that is not functionally required, and would also require them to configure privacy settings by default to the most privacy-protecting options. Additionally, companies would be required to conduct assessments of whether their products and services, such as content, algorithms, and targeted advertising systems, may be harmful to children.[90] Modeled after CAADCA, the Maryland Age-Appropriate Design Code, also known as the "Kids Code," was signed into law in 2024. There are also similar efforts at the federal level, with the Kids Online Safety Act and COPPA 2.0, together known as "KOSPA," which aim to increase privacy protections and mitigate harms for children online.[91]

---

[87] *Complying with COPPA: Frequently Asked Questions*, *supra* note 82.

[88] *Introduction to the Children's code*, INFO. COMM'R OFF., https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/childrens-information/childrens-code-guidance-and-resources/introduction-to-the-childrens-code/ [https://perma.cc/ET63-KJB9] (last visited Apr. 28, 2025).

[89] Chloe Altieri & Kewa Jiang, *California Age-Appropriate Design Code Aims To Address Growing Concern About Children's Online Privacy And Safety*, FUTURE PRIV. F. (June 28, 2022) https://fpf.org/blog/california-age-appropriate-design-code-aims-to-address-growing-concern-about-childrens-online-privacy-and-safety/ [https://perma.cc/E6QK-LV8G].

[90] CAL. CIV. CODE § 1798.99.31 (2025).

[91] Caitlin Burke et al., *KOSA's Path Forward: Distinguishing Design from Content to Maintain Free Speech Protections*, TECH POL'Y PRESS (Oct. 1, 2024), https://www.techpolicy.press/kosas-path-forward-distinguishing-design-from-content-to-maintain-free-speech-protections/ [https://perma.cc/4ZAQ-XH2B]; Bailey Sanchez & Nick Alereza, *Contextualizing The Kids Online Safety And Privacy Act: A Deep Dive Into*

Despite these efforts, CAADCA, KOSPA, and the like are facing significant legal challenges largely related to concerns regarding freedom of speech under the First Amendment, as shown by the complaint filed by NetChoice in 2022 regarding CAADCA.[92] NetChoice is a trade association of businesses, including companies such as Amazon, Google, Meta, Netflix, and others.[93] A coalition of 19 states and Washington, D.C., in support of CAADCA, argued that invalidating it and similar laws could "strangle state efforts to regulate harmful internet practices in their infancy."[94] Their effort was joined by Common Sense Media, a nonprofit organization dedicated to protecting children online and increasing digital literacy, which also issued statements in support of CAADCA and KOSPA.[95] Additionally, the FTC announced changes to the COPPA Rule in January 2025 with the aim of limiting monetization of children's personal information, such as requiring opt-in parental consent for third-party advertising.[96]

The Federal Kids Bill, FUTURE PRIV. F. (Aug. 2, 2024) https://fpf.org/blog/contextualizing-the-kids-online-safety-and-privacy-act-a-deep-dive-into-the-federal-kids-bill/ [https://perma.cc/B7E3-X4WJ].

[92] Complaint, NetChoice v. Bonta, No. 5:24-cv-07885-EJD, 2024 WL 5264045, (N.D. Cal. Dec. 31, 2022) (Doc. 1).

[93] *About Us*, NETCHOICE, https://netchoice.org/about/#our-mission [https://perma.cc/LT36-H3BZ] (last visited Apr. 28, 2025).

[94] Gabby Miller, *Maryland Kids Code Signed Into Law, But May Face Legal Challenges*, TECH POL'Y PRESS (May 10, 2024), https://www.techpolicy.press/maryland-kids-code-becomes-law/ [https://perma.cc/92M9-BTBX]; Brief for Nevada et al. as Amici Curiae Supporting Defendant-Appellant, Netchoice v. Bonta, 113 F.4th 1101 (9th Cir. 2024) (No. 23-2969).

[95] *Statement on Lawsuit Filed by Lobbyists for Big Tech Seeking to Block the California Age-Appropriate Design Code Act*, COMMON SENSE (Dec. 15, 2022) https://www.commonsensemedia.org/press-releases/statement-on-lawsuit-filed-by-lobbyists-for-big-tech-seeking-to-block-the-california-age-appropriate [https://perma.cc/6X3T-CP7Z]; *KOSA One Pager*, COMMON SENSE, https://www.commonsensemedia.org/sites/default/files/featured-content/files/kosa-one-pager.pdf [https://perma.cc/V93Q-83CG]; *COPPA 2.0 One Pager*, COMMON SENSE, https://www.commonsensemedia.org/sites/default/files/featured-content/files/coppa_2.0_one_pager_2021.pdf [https://perma.cc/6S32-5P3H] (last visited Apr. 28, 2025).

[96] Press Release, FTC, FTC Finalizes Changes to Children's Privacy Rule Limiting Companies' Ability to Monetize Kids' Data (Jan. 16, 2025), https://www.ftc.gov/news-events/news/press-releases/2025/01/ftc-finalizes-

In summary, this case study regarding companies' misrepresentations of how they protect children's online privacy as well as their misuse of children's data for targeted content and advertising purposes illustrates their hypocrisy in data-governance practices. Companies leverage their internal capabilities to identify and target children with advertisements while claiming to lack "actual knowledge" of those users' ages, as required by COPPA. They know there are children using their services and yet do not fully respond to their vulnerabilities through safeguards or more protective design. Companies also continue to push back on proposed legislation that would hinder their business interests regarding children on their platforms. Thus, companies continue to publicly boast about their commitment to protect children online, while actively failing to do so.

II. The Case of Identifying Health Information

The Health Insurance Portability and Accountability Act of 1996 (HIPAA) is the federal law that protects health information shared within the context of healthcare to providers and their business associates.[97] The Department of Health and Human Services (HHS) issued the HIPAA Privacy Rule to implement HIPAA privacy requirements in 2000.[98] The HIPAA Privacy Rule establishes safeguards to protect the privacy of protected health information (PHI).[99] Specifically,

---

changes-childrens-privacy-rule-limiting-companies-ability-monetize-kids-data [https://perma.cc/VP9K-LFTQ].

[97] Health Insurance Portability and Accountability Act of 1996, codified at 42 U.S.C. § 1320d *et seq.* (2024).

[98] Standards for Privacy of Individually Identifiable Health Information, 45 C.F.R. § 160, 164 (2025). HIPAA Privacy Rule has subsequently been updated several times to refine its application and enforcement. *The HIPAA Privacy Rule*, U.S. DEP'T OF HEALTH & HUM. SERVS., https://www.hhs.gov/hipaa/for-professionals/privacy/index.html [https://perma.cc/8HWT-VSPU] (last visited Apr. 28, 2025).

[99] PHI is defined as "individually identifiable health information" (IIHI) that relates to "past, present, or future physical or mental health or condition of an individual; the provision of health care to an individual; or the past, present, or future payment for the provision of health care to an individual" and either "identifies the individual" or "can [reasonably] be used to identify the individual." *See* 45 C.F.R. § 160.103, https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-C/part-

the Rule requires an individual's authorization outside of routine care, payment, and administrative operations before PHI can be used or disclosed by "covered entities" such as healthcare providers and their "business associates" such as claims processors.[100] HIPAA also guarantees the data subject rights with respect to their PHI, such as the right to access or amend their information.[101]

Patients increasingly rely on websites and mobile apps offered by healthcare providers to access medical services, communicate with their providers, and manage their medical records.[102] These interfaces commonly use third-party tracking tools for analytics and advertising.[103] These third-party tracking tools can be used to monitor when and how a patient navigates a website or mobile app, profile a patient's health conditions or interests, and subsequently run targeted advertising campaigns

---

160/subpart-A/section-160.103 [https://perma.cc/BQ49-9F6L]. Also note that IIHI is not PHI under HIPAA if it is created or received by HIPAA covered entities such as health care providers. *See What is Considered PHI Under HIPAA?*, The HIPAA Journal, https://www.hipaajournal.com/considered-phi-hipaa/ [https://perma.cc/WZY3-NVS5] (last visited Apr. 28, 2025).

[100] PHI is defined as "individually identifiable health information" (IIHI) that relates to "past, present, or future physical or mental health or condition of an individual; the provision of health care to an individual; or the past, present, or future payment for the provision of health care to an individual" and either "identifies the individual" or "can [reasonably] be used to identify the individual." *See* 45 C.F.R. 160.103. Also note that IIHI is not PHI under HIPAA if it is created or received by HIPAA covered entities such as health care providers. *See What is Considered PHI Under HIPAA?*, HIPAA J., https://www.hipaajournal.com/considered-phi-hipaa/ [https://perma.cc/WZY3-NVS5] (last visited Apr. 28, 2025).

Summary of the HIPAA Privacy Rule, U.S. DEP'T OF HEALTH & HUM. SERVS., https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html [https://perma.cc/3FZK-XW48] (last visited Apr. 28, 2025) (requiring authorization for uses and disclosures of personal health information that is not for treatment, payment or health care operations).

[101] *Id.*

[102] Wesley Barker, et al., *The Evolution of Health Information Technology for Enhanced Patient-Centric Care in the United States: Data-Driven Descriptive Study*, 26 J. MED. INTERNET RSCH. e59791 (2024); Wesley Barker & Chelsea Richwine, *Patient usage of apps to access online medical records*, 6 JAMA NETWORK OPEN, no. 11 (2023).

[103] Ari B. Friedman et al., *Widespread Third-Party Tracking On Hospital Websites Poses Privacy Risks For Patients And Legal Liability For Hospitals*, 42 HEALTH AFFS. 508 (2023).

to recruit new patients and engage existing patients with specific health conditions or interests.[104] Like other industries, online advertising now accounts for the vast majority of the total ad spend for the healthcare industry.[105] The online ad spend by the U.S. healthcare and pharmaceutical industry is projected to exceed $30 billion in 2024.[106]

Healthcare provider websites and apps install third-party advertising and tracking tools to personalize and optimize online ad campaigns.[107] When installed, third-party advertising and tracking tools can collect PHI when a patient searches for specific medical conditions, schedules an appointment with a health care provider, pays their bill, submits a health risk assessment, or accesses medical records via health care provider websites and apps. The HIPAA Privacy Rule prohibits disclosure and use of such PHI to vendors of these

---

[104] Kevin D. StClerg, *Digital Marketing for Private Practice: How to Attract New Patients. In Seminars in Hearing*, 40 THIEME MED. PUBLISHERS 260 (2019); *Expanding your reach through consumer marketing,* OPTUM, https://cdn-aem.optum.com/content/dam/optum3/optum/en/resources/case-studies/cas-expanding-your-reach-through-consumer-marketing-case-study.pdf [https://perma.cc/LEN8-6TGJ] (last visited Apr. 29, 2025); *Providence rebuilds Google Ads campaigns to reach more patients*, PROVIDENCE, https://www.wpromote.com/wp-content/uploads/2020/06/GPAR-SCS-1830_Wpromote-Providence.pdf [https://perma.cc/4DBX-AQCR] (last visited Apr. 29, 2025); *Healthcare marketing case study: 27x more patient bookings*, HEALTHTECH HIPPO, https://healthtechhippo.com/healthcare-marketing-case-study/ [https://perma.cc/4W8M-4NBP].

[105] Rajiv Leventhal, *US Healthcare and Pharma Ad Spending 2024*, EMARKETER (Oct. 25, 2024), https://www.emarketer.com/content/us-healthcare-pharma-ad-spending-2024 [https://perma.cc/B57D-5K93].

[106] Rajiv Leventhal, *Pharma accounts for nearly 90% of the broader industry's digital ad spending*, EMARKETER (Oct. 20, 2024), https://www.emarketer.com/content/pharma-accounts-nearly-90-of-broader-industry-digital-ad-spending [https://perma.cc/JS8Y-T7H6].

[107] Several major tech companies provide such services. *See, e.g.*, *Meta Pixel*, META, https://www.facebook.com/business/tools/meta-pixel [https://perma.cc/TF66-T5UA] (last visited Mar. 12, 2026) ("Add a piece of code to your website that lets you measure, optimize and build audiences for your ad campaigns."); *Google Analytics Best Practices*, GOOGLE ADS https://support.google.com/google-ads/answer/6175315 [https://perma.cc/QPH7-UWET] (last visited Mar. 12, 2026); *About TikTok Pixel*, TIKTOK https://ads.tiktok.com/help/article/tiktok-pixel [https://perma.cc/W4SF-3ZE8] (last visited Mar. 12, 2026).

third-party tracking tools unless they have signed a business associates agreement (BAA) that requires them to comply with HIPAA and its Privacy Rule.[108] However, the vendors of the most commonly used advertising and tracking tools do not sign BAAs or ask advertisers to not share information that may be covered under HIPAA.[109]

---

[108] *Sample Business Associate Contracts*, U.S. DEP'T OF HEALTH & HUM. SERVS., https://www.hhs.gov/hipaa/for-professionals/covered-entities/sample-business-associate-agreement-provisions/index.html [https://perma.cc/PUY4-R4HK] (last visited Apr. 29, 2025). Note that simply informing patients about the use of tracking technologies in the privacy policy, notice, or terms and conditions is not sufficient for HIPAA compliance. *See Use of Online Tracking Technologies by HIPAA Covered Entities and Business Associates*, U.S. DEP'T OF HEALTH & HUM. SERVS., https://www.hhs.gov/hipaa/for-professionals/privacy/guidance/hipaa-online-tracking/index.html#ftn36 [https://perma.cc/F4UQ-HF6L] (last visited Apr. 29, 2024).

[109] *HIPAA and Google Analytics*, GOOGLE, https://support.google.com/analytics/answer/13297105?hl=en [https://perma.cc/H69Y-8R85] (last visited Apr. 29, 2025) ("Customers must refrain from using Google Analytics in any way that may create obligations under HIPAA for Google. HIPAA-regulated entities using Google Analytics must refrain from exposing to Google any data that may be considered Protected Health Information (PHI), even if not expressly described as PII in Google's contracts and policies. Google makes no representations that Google Analytics satisfies HIPAA requirements and does not offer Business Associate Agreements in connection with this service."); *About prohibited information*, META, https://www.facebook.com/business/help/361948878201809 [https://perma.cc/FB4D-868F] (last visited Apr. 29, 2025) ("You must not share any of the following with Meta via the Meta Business Tools" . . . . Information about an individual's physical or mental health, such as . . . 'Diseases, medical conditions and injuries', 'Sexual and reproductive health', 'Mental health and psychological states', 'Medical procedures, treatments and testing', 'Medications and supplements (OTC and prescription)', 'Physical locations that identify a health condition, or places of treatment and counseling', etc."). There exist tracking tools that are specifically advertised as HIPAA compliant and sign BAA. *See*, *e.g.*, *HIPAA-compliant analytics and data activation: Unlock marketing insights with Piwik PRO*, PIWIK, https://piwik.pro/hipaa/ [https://perma.cc/76VS-4FLY] (last visited Apr. 29, 2025) *How Mixpanel provides HIPAA-compliant analytics for healthcare*, MIXPANEL, https://mixpanel.com/blog/hipaa-compliant-analytics-healthtech/ [https://perma.cc/GX2V-VS38](last visited Apr. 29, 2025); *PostHog & HIPAA*, POSTHOG, https://posthog.com/docs/privacy/hipaa-compliance [https://perma.cc/4F5J-2HNH] (last visited Apr. 29, 2025).

Although popular advertising and tracking tools such as Google Analytics[110] and Meta Pixel[111] are not designed to be HIPAA compliant, healthcare provider websites and apps have continued using them, potentially disclosing PHI that is covered under HIPAA.[112] Despite possessing the capability to determine specific medical conditions and interests of individuals for targeted advertising (in line with industry standards such as those outlined in the Interactive Advertising Bureau (IAB) content taxonomy), vendors of these advertising and tracking tools do not implement sufficient safeguards to detect and prevent the potential collection of HIPAA-covered PHI.[113]

In this case study, we make a case that while tech companies have the capability to target individuals based on their health-related profiles, they do not leverage the same capability to ensure that their tracking tools do not collect or use PHI received from HIPAA-covered entities such as healthcare providers for targeted advertising. This reflects another instance of hypocrisy in data governance: tech companies use their capability for advertising when it serves their business interests while neglecting it in protecting consumer/patient privacy.

Tech companies have developed sophisticated capabilities to detect and profile health information for targeted advertising. The IAB's standardized content taxonomy includes potentially sensitive medical segments such as "Birth Control" and "Substance Abuse" that enable advertisers to target users based on specific health-related interests.[114]

---

[110] *Start learning about Google Analytics*, GOOGLE, https://developers.google.com/analytics [https://perma.cc/E7UX-86EV] (last visited Apr. 29, 2025).

[111] *Meta Pixel*, *supra* note 107.

[112] *Facebook is Receiving Sensitive Medical Information from Hospital Websites*, MARKUP (June 16, 2022), https://themarkup.org/pixel-hunt/2022/06/16/facebook-is-receiving-sensitive-medical-information-from-hospital-websites [https://perma.cc/UGL2-4QCB]; Friedman et al., *supra* note 103.

[113] *Content Taxonomy*, INTERACTIVE ADVERT. BUREAU, https://www.iab.com/guidelines/content-taxonomy/ [https://perma.cc/2F2H-JBCT] (last visited Apr. 29, 2025).

[114] *Taxonomies*, GITHUB, https://github.com/InteractiveAdvertisingBureau/Taxonomies/blob/develo

Google Ads, for example, provides advertisers with a taxonomy that allows them to target ad campaigns to individuals based on highly specific medical segments such as "Mental Health," "Eating Disorders," and "AIDS & HIV."[115] Facebook also allows advertisers to target ads to individuals based on specific behaviors, such as searching for certain medical conditions online.[116] For example, Facebook's "Detailed Targeting" program uses third-party data to allow advertisers to target women who purchase plus-size clothing with bariatric-surgery ads.[117] In addition, both Google and Facebook allow advertisers to "retarget"[118] users who have taken a specific action (such as a patient who booked an appointment), target users who are likely to take a specific action,[119] or target "lookalikes" of users who have taken a specific action.[120] Lookalikes here refers to new users who are determined to share characteristics with or "look like" an existing set of users.[121]

---

p/Content%20Taxonomies/Content%20Taxonomy%203.1.tsv [https://perma.cc/K9A3-PTDR] (last visited Apr. 29, 2025).

[115] *Topics*, GOOGLE, https://developers.google.com/google-ads/api/data/topics [https://perma.cc/D4T3-62J6] (last visited Apr. 29, 2025).

[116] *Facebook Ad Targeting for Healthcare Marketers*, WEBMD IGNITE, https://webmdignite.com/blog/facebook-ad-targeting-healthcare-marketers [https://perma.cc/PLK8-DESL] (last visited Apr. 29, 2025).

[117] *Id.*

[118] *Retargeting*, GOOGLE, https://developers.google.com/tag-platform/devguides/remarketing [https://perma.cc/2UXA-BRZG] (last visited Apr. 29, 2025); *Retargeting*, META, https://www.facebook.com/business/goals/retargeting [https://perma.cc/M6T8-7M5L] (last visited Apr. 29, 2025).

[119] *About Target CPA bidding*, GOOGLE, https://support.google.com/google-ads/answer/6268632 [https://perma.cc/8GJ2-VCVU] (last visited Apr. 29, 2025); *Troubleshoot conversion optimization*, META, https://www.facebook.com/business/help/197634954160445 [https://perma.cc/F7U4-49P6] (last visited Apr. 29, 2025).

[120] *Use Lookalike segments to grow your audience*, GOOGLE, https://support.google.com/google-ads/answer/13541369?hl=en [https://perma.cc/TAR7-ZUVE] (last visited Apr. 29, 2025).

[121] *About lookalike audiences*, Meta, https://www.facebook.com/business/help/164749007013531 [https://perma.cc/5KTH-GJDQ] (last visited Apr. 29, 2025).

A number of data brokers also make available lists of users that correspond to specific health conditions such as asthma, psoriasis, chronic idiopathic constipation, and rheumatoid arthritis;[122] Facebook users who are "likely to need bariatric surgery, total knee or hip replacement, stroke, A-Fib or breast cancer";[123] users who visit pages about specific types of cancer,[124] such as bladder cancer,[125] colon cancer,[126] and Hodgkin's lymphoma;[127] and users who are interested in "medicine & drugs."[128] According to a 2023 survey of data brokers, nearly a dozen were willing and able to sell lists of individuals—that include Personally Identifiable Information (PII) such as names and addresses—associated with conditions such as depression, anxiety, and bipolar disorder.[129]

There is ample evidence that online advertising and tracking tools are collecting potentially HIPAA-covered PHI from healthcare-provider websites and apps. 90% of health-

---

[122] *Welcome to the DeepIntent Audience Marketplace*, DEEPINTENT, https://web.archive.org/web/20250806163213/https://www.deepintent.com/wp-content/uploads/2022/03/2022_DeepIntent_Partner-Sell-Sheet_Audience-Marketplace_ShareThis.pdf (last visited Apr. 29, 2025).

[123] *Expanding your reach through consumer marketing,* OPTUM, https://cdn-aem.optum.com/content/dam/optum3/optum/en/resources/case-studies/cas-expanding-your-reach-through-consumer-marketing-case-study.pdf [https://perma.cc/LEN8-6TGJ] (last visited Apr. 29, 2025).

[124] *Cancer Targeting*, HEALTHYADS, https://www.healthyads.com/targeting/cancer-targeting/ [https://perma.cc/4VLG-FSLS] (last visited Apr. 29, 2025).

[125] *Bladder Cancer Targeting*, HEALTHYADS, https://www.healthyads.com/targeting/cancer-targeting/bladder-cancer-targeting/ [https://perma.cc/Y86U-WERQ] (last visited Apr. 29, 2025).

[126] *Colon Cancer Targeting*, HEALTHYADS, https://www.healthyads.com/targeting/cancer-targeting/colon-cancer-targeting/ [https://perma.cc/W77R-QSLG] (last visited Apr. 29, 2025).

[127] *Hodgkin's Lymphoma Targeting*, HEALTHYADS, https://www.healthyads.com/targeting/cancer-targeting/hodgkins-lymphoma-targeting/ [https://perma.cc/L2SZ-JE4Y] (last visited Apr. 29, 2025).

[128] *2019 Data Directory,* ORACLE DATA CLOUD (2019), https://www.oracle.com/us/solutions/cloud/data-directory-2810741.pdf [https://perma.cc/6GJD-3WGH].

[129] Joanne Kim, *Data brokers and the sale of Americans' mental health data*, DUKE SANFORD SCH. OF PUB. POL'Y (Feb. 2023), https://techpolicy.sanford.duke.edu/wp-content/uploads/2023/02/Kim-2023-Data-Brokers-and-the-Sale-of-Americans-Mental-Health-Data.pdf [https://perma.cc/QK2D-8QVP].

related webpages disclose information about specific symptoms, treatments, and diseases to third parties such as Google and Facebook.[130] In one study, all health-related websites examined disclosed information to third parties, with Google and Facebook being the most common parties to which information was disclosed.[131] 14% of the patient portals include third-party tracking tools, with most significant PHI disclosures happening to Google and Facebook.[132] Third-party tracking tools were present on 98.6% of hospital websites, with 98.5% disclosing information to Google and 55.6% to Facebook.[133]

In 2022, *Markup* investigated the top one hundred U.S. hospitals and found that thirty-three of them had installed Meta Pixel.[134] They found that disclosures to Facebook included the doctor's name, selected conditions, and the search terms used when a patient schedules an appointment. They also found disclosures to Facebook inside patient portals, including medication names, allergic reactions, and appointment details.[135] Later in 2022, the Office for Civil Rights (OCR) in HHS issued a Bulletin[136] to warn health care providers about potential violations of the HIPAA Privacy Rule due to the use of tracking technologies on their websites and patient portals.[137] The Bulletin specified that the use of tracking tools

---

[130] Timothy Libert, *Privacy Implications of Health Information Seeking on the Web*, 58 COMMC'NS ACM 68 (2015).

[131] Alexander R. Zheutlin et al., *Data-Tracking on Government, Non-profit, and Commercial Health-Related Websites*, 37 J. GEN. INTERNAL MED. 1315 (2021).

[132] Mingjia Huo et al., *All Eyes on Me: Inside Third Party Trackers' Exfiltration of PHI from Healthcare Providers' Online Systems*, 21 PROCS. WORKSHOP ON PRIV. ELEC. SOC'Y 197 (2022).

[133] Ari B. Friedman et al., *supra* note 103.

[134] *Facebook is Receiving Sensitive Medical Information from Hospital Websites*, *supra* note 112.

[135] *Id.*

[136] A portion of the bulletin about disclosure of IP address on an unauthenticated webpage addressing specific health conditions or health care providers was vacated. Am. Hosp. Ass'n v. Becerra, 4:23-cv-01110-P (N.D. Tex. June 20, 2024); *Use of Online Tracking Technologies by HIPAA Covered Entities and Business Associates*, *supra* note 108.

[137] Steve Alder, *OCR, FTC Publish Online Tracking Technology Warning Letters*, HIPAA J. (Sep. 6, 2023), https://www.hipaajournal.com/ocr-ftc-publish-online-tracking-technology-warning-letters/ [https://perma.cc/6NGD-XXYE].

on user-authenticated webpages, which require a user to log in before they are able to access the webpage, may result in an impermissible disclosure of PHI under the HIPAA Privacy Rule, unless the tracking technology vendor enters into a BAA with a commitment to comply with the HIPAA Privacy Rule. The Bulletin further discussed the use of tracking technologies on unauthenticated web pages (for instance, the login page of a patient portal, or webpages that address specific symptoms or health conditions). In July 2023, the FTC and HHS issued warning letters to 130 entities, including healthcare providers, about the use of tracking technologies in a manner that would result in impermissible disclosures of PHI to third parties, citing academic research and the *Markup*'s investigation.[138]

Tech companies that provide tracking technologies without committing to HIPAA compliance through signing a BAA when HIPAA-covered entities use their tracking technology can leverage their existing capabilities to identify and use health information in advertising for HIPAA compliance. As outlined below, the following two-step process can be used to determine whether a tech company's tracking technology is receiving potentially HIPAA-covered PHI: first, identify if the website or app is associated with a HIPAA-covered entity; and second, assess whether the use of a tracking tool on a webpage may result in disclosure of PHI.

First, to figure out whether the website is associated with a HIPAA-covered entity such as a healthcare provider, companies can rely on the simple fact that the HIPAA Privacy Rule requires covered entities to provide a "Notice of Privacy Practices" whose content is required to include specific language.[139] The websites of covered entities typically include a

---

[138] Letters from Melanie Fontes Rainer, Dir., Off. of C.R., U.S. Dep't of Health & Hum. Servs., and Samuel Levine, Dir., Bureau of Cons. Prot., FTC, on Online Tracking Technologies (July 20, 2024), https://www.hhs.gov/sites/default/files/ocr-ftc-letters-re-use-online-tracking-technologies.pdf [https://perma.cc/7LTY-MU6S]; Press Release, FTC, FTC and HHS Warn Hospital Systems and Telehealth Providers about Privacy and Security Risks from Online Tracking Technologies (July 20, 2023), https://www.ftc.gov/news-events/news/press-releases/2023/07/ftc-hhs-warn-hospital-systems-telehealth-providers-about-privacy-security-risks-online-tracking [https://perma.cc/35H4-WJVB].

[139] The notice must contain the following statement as a header or otherwise prominently displayed: "THIS NOTICE DESCRIBES HOW MEDICAL

link to a webpage titled "Notice of Privacy Practices," "HIPAA Privacy Practices," or "HIPAA Notice of Privacy Rights" on their homepage indicating that the website is associated with a HIPAA-covered entity.[140] Tech companies can automatically crawl the websites where their advertising and tracking tools are deployed to detect the presence of the HIPAA privacy notice for determining whether the website is associated with a HIPAA-covered entity. In fact, tech companies routinely crawl websites at scale.[141] Thus, tech companies can analyze crawl data for the websites where their advertising and tracking tools are installed and then simply search for specific phrases in the webpage's content such as "Notice of Privacy Practices," "HIPAA Privacy Practices," or "HIPAA Notice of Privacy Rights," which are indicative of the fact that the website is associated with a HIPAA-covered entity.

Second, to figure out whether a specific webpage where tracking tools are deployed may result in disclosure of PHI, companies can rely on the guidance provided in the Bulletin published by HHS.[142] The Bulletin distinguishes between two types of webpages on a HIPAA-covered entity's website: "user-authenticated" (that require users to log in before accessing it) and "unauthenticated" (that are publicly accessible without requiring a user to log in):

1. For user-authenticated webpages of HIPAA covered entities, the Bulletin states that "user-authenticated

---

INFORMATION ABOUT YOU MAY BE USED AND DISCLOSED AND HOW YOU CAN GET ACCESS TO THIS INFORMATION. PLEASE REVIEW IT CAREFULLY." *See* 45 C.F.R. § 164.520 (2025).

[140] For example, see "Notice of Privacy Practices" on the websites of Mayo Clinic (https://www.mayoclinic.org), Cleveland Clinic (https://my.clevelandclinic.org), and the Johns Hopkins Hospital (https://www.hopkinsmedicine.org).

[141] *Overview of Google Crawlers and Fetchers (User Agents)*, GOOGLE SEARCH CENT., https://developers.google.com/crawling/docs/crawlers-fetchers/overview-google-crawlers [https://perma.cc/6S4J-KX2Q]; *Meta Web Crawlers*, META FOR DEVS., https://developers.facebook.com/docs/sharing/webmasters/web-crawlers/ [https://perma.cc/P9BU-SN6Q].

[142] *Use of Online Tracking Technologies by HIPAA Covered Entities and Business Associates*, *supra* note 108.

web pages generally have access to PHI" and that they should "only use and disclose PHI in compliance with the HIPAA Privacy Rule."[143] Put simply, all user-authenticated webpages of a HIPAA-covered entity are likely subject to HIPAA as per the guidance in the Bulletin.

2. For unauthenticated webpages of HIPAA-covered entities, the Bulletin states that "HIPAA Rules apply" "in some cases" that include:
    a. Webpages where users can "register" or "login";
    b. Webpages where users can "search for doctors or schedule appointments"; and
    c. Webpages that address "specific symptoms or health conditions."[144]

Automatically detecting user-authenticated webpages can be straightforward. Webpages typically return HTTP 4XX status codes when a user attempts to access a webpage without requisite credentials. To detect such webpages, one can rely on whether accessing a given webpage without requisite credentials results in an HTTP 4XX status code such as 401 Unauthorized or 403 Forbidden.[145] Others redirect to the login pages using HTTP 3XX status codes when an individual attempts to access a webpage without the requisite authentication. To detect such websites, one can rely on whether accessing a given webpage results in an HTTP 3XX status code such as 302 Found Redirect or 307 Temporary Redirect.[146]

All other webpages are considered unauthenticated and can be classified into one of the categories below to determine whether they are potentially subject to HIPAA under the

---

[143] *See The HIPAA Privacy Rule*, *supra* note 98.
[144] *Id.*
[145] *HTTP response status codes*, MDN, https://developer.mozilla.org/en-US/docs/Web/HTTP/Reference/Status#client_error_responses [https://perma.cc/D7X8-5TWD] (last visited Jan. 29, 2026).
[146] *Id.*

Bulletin. Tech companies can first crawl the unauthenticated webpages of the website of a HIPAA-covered entity where their advertising and tracking tools are deployed using their existing crawlers. Tech companies can then automatically detect whether the information on an unauthenticated webpage may contain PHI that is subject to HIPAA as follows.

1. Webpages where users can "register" or "login" can be detected using (a) regular expressions (for keywords such as "register" or "login") in the URL or page content, (b) HTML-form analysis to search for tag types such as "email" or "password", or (c) the browser's Autofill feature;
2. Webpages where users can "search for doctors or schedule appointments" can be detected using (a) regular expressions (for phrases such as "Book an Appointment," "Schedule a Visit," "Find a Doctor") in the URL or page content or (b) HTML-form analysis to search for tag types "submit" with text such as "Book Appointment";
3. Webpages that address "specific symptoms or health conditions" can be detected by searching the webpage content using the MetaMap application developed by the National Library of Medicine (NLM) to identify Unified Medical Language System (UMLS) concepts that comprehensively cover diagnosis, procedures, diseases, drugs, etc.[147]

Recent advances in generative artificial intelligence (GenAI), particularly the large language models (LLMs) offered by tech companies like Google, Meta, and OpenAI, further strengthen the feasibility of this approach. With these

---

[147] *MetaMap*, GITHUB, https://www.nlm.nih.gov/research/umls/implementation_resources/metamap.html [https://perma.cc/F5C7-4ETH] (last visited Apr. 29, 2025); *UMLS Metathesaurus Vocabulary Documentation*, NAT'L INST. OF HEALTH, https://www.nlm.nih.gov/research/umls/sourcereleasedocs/index.html [https://perma.cc/B4FX-2K35] (last visited Apr. 29, 2025).

powerful GenAI models, tech companies can move beyond simple keyword searches to classify whether a website belongs to a HIPAA-covered entity and whether a specific page may contain PHI. These models already meet or surpass state-of-the-art accuracy on a wide range of tasks,[148] including health-related tasks.[149] Tech companies are already leveraging these GenAI models in advertising with great success.[150] They thus can likely also be readily used to identify collection of potentially HIPAA-covered data.

In summary, this case study about potential HIPAA privacy violations via advertising and tracking tools on healthcare provider websites and apps illustrates another hypocrisy in data-governance practices of tech companies. Tech companies readily leverage their capabilities for targeted advertising to users based on health information, yet they fail to commit to protect health data by applying the same set of capabilities.

Conclusion

These case studies provide concrete illustrations of the hypocrisy in data governance. Tech companies navigate the ambiguities of data governance, both as a concept and with regard to their specific data-governance capabilities, to their benefit. In doing so, they get to have it both ways. Companies

---

[148] Yazhou Zhang, et al., *Pushing the Limit of LLM Capacity for Text Classification*, 2025 COMPANION PROCS. ACM ON WEB CONF. 1524.

[149] Rahul K. Arora, et al., *HealthBench: Evaluating Large Language Models Towards Improved Human Health* (May 13, 2025) (unpublished manuscript), https://cdn.openai.com/pdf/bd7a39d5-9e9f-47b3-903c-8b847ca650c7/healthbench_paper.pdf [https://perma.cc/W3WD-KAE9].

[150] Sheila Dang, *Google upgrades AI product for advertisers with Gemini models*, REUTERS (Feb. 27, 2024), https://www.reuters.com/technology/google-upgrades-ai-product-advertisers-with-gemini-models-2024-02-22/ [https://perma.cc/V3E9-NLYS] ("Google said Gemini will improve a product called Performance Max, which automatically finds the best placements for a brand's ads across Google services including email, search and YouTube."); *Generative AI features for ads coming to all advertisers*, META (Oct. 4, 2023), https://www.facebook.com/business/news/generative-ai-features-for-ads-coming-to-all-advertisers [https://perma.cc/G7ZP-NEYJ]; Jordan Marlow, *How AI Has Boosted Meta's Advertising Business*, IDX (May 24, 2024), https://www.idx.inc/newsroom/how-ai-has-boosted-metas-advertising-business [https://perma.cc/J9VG-63NP].

develop sophisticated techniques to identify children on their platforms or obtain commercially valuable health data, yet they claim ignorance of that very same data when facing legal scrutiny. We must demand increased clarity and consistency from tech companies.

COPPA, HIPAA, and their attendant rules were written at a time when companies' ability to draw reliable inferences from their data was far less sophisticated. At the time, it was likely quite difficult for companies to identify young children through their browsing activity or determine websites' HIPAA-coverage through web-scraping techniques. In this context, it perhaps made sense to let companies use ignorance to disclaim their data-processing obligations. Now, however, such inferences are not only possible and cost-efficient; they are a key revenue stream as well. Tech companies' business models reflect that change in technology, and so too should their legal obligations.

In addition to the technical solutions described earlier, we also propose two policy interventions to help mitigate tech companies' hypocrisy regarding data governance. First, we argue that judges should be more willing to take into account tech companies' actual data-governance capabilities, as promoted or demonstrated in other contexts, when applying the scope of their data stewardship obligations under existing laws. Second, we argue that policymakers should change the law to expand the scope of data stewardship obligations in light of tech firms' sophisticated capabilities. We elaborate on each of the two policy approaches next.

First, we propose that **regulators and courts more explicitly account for demonstrated data-governance capabilities** (and failures) when evaluating compliance, reasonableness, and culpability. That implies more exacting scrutiny of unsupported assertions about what a company "could not" know or control, particularly where its systems are designed to infer user attributes or segment audiences. For example, COPPA uses an "actual knowledge" standard to trigger certain responsibilities.[151] The FTC has clarified that "actual

[151] *When does the operator of a website or online service have "actual knowledge" of someone's age?,* FED. TRADE COMM'N,

knowledge" under COPPA includes cases of "willful disregard"—situations where platforms "blind themselves" to clear evidence of a user's age.[152] However, COPPA itself does not explicitly define "willful disregard," and recent FTC rulemakings have declined to expand the standard so far as to include "constructive knowledge."[153] Other privacy laws have likewise hinged on this literal interpretation of "actual knowledge" of a data's heightened sensitivity.[154] Likewise, data controllers avoid complying with HIPAA's data protections by claiming they are unaware that they are receiving protected information and failing to complete BAAs.[155] Yet, as we have shown, they have any number of technically feasible methods of identifying potentially HIPAA-covered data and preventing unlawful disclosures.

The obvious problem with strictly interpreting knowledge requirements is that it incentivizes willful ignorance. Tech companies are rewarded for avoiding direct knowledge that the data they are collecting is subject to heightened protection. Meanwhile, they continue to develop increasingly sophisticated data-governance techniques. These techniques allow companies to make reliable inferences about data

---

https://www.ftc.gov/business-guidance/resources/childrens-online-privacy-protection-rule-not-just-kids-sites#when [https://perma.cc/8EWM-C7ND].

[152] *Willful Disregard Under the CPA: How Do Online Services Know Users Are Minors?*, GEORGETOWN L. TECH. INST., https://www.law.georgetown.edu/tech-institute/wp-content/uploads/sites/42/2025/09/Willful-Disregard-Knowledge-Standard-CPA.pdf [https://perma.cc/CC2W-TP5G] (citing 89 Fed. Reg. 2034, 2037 (2024) ("The concept of actual knowledge includes willful disregard. *See, e.g.*, Glob.-Tech Appliances, Inc. v. SEB S.A., 563 U.S. 754, 766 (2011) (noting that '[i]t is also said that persons who know enough to blind themselves to direct proof of critical facts in effect have actual knowledge of those facts'). Therefore, the Rule already applies to instances in which an operator of a general audience site or service willfully disregards the fact that a particular user is a child."); Camille Altieri, *Now, on the Internet, Will Everyone Know If You're a Child?*, FUTURE PRIV. F., 2024, https://fpf.org/blog/now-on-the-internet-will-everyone-know-if-youre-a-child/ [https://perma.cc/PD33-T5AP] (describing "constructive knowledge" as a "should have known" standard).

[153] *Id.*

[154] *Id.*

[155] *See supra* Part II.

without needing the user to directly reveal that information.[156] This reflects a deeper observation about how these tech companies operate. They make decisions not just based on information about which they are certain but based on highly sophisticated and educated guesses.[157] In other words, tech companies are getting increasingly good at obtaining and profiting off of actionable insights without ever receiving "actual knowledge" in a literal sense.

The gap between actionable insight and legally sufficient actual knowledge must be bridged. To do so, judges should look beyond tech companies' feigned ignorance. Rather than relying on a literal interpretation of "actual knowledge," judges should interpret the term to include predictive inferences with a sufficiently high confidence interval. For example, if a company is confident enough that a particular cohort of users are children to make internal decisions based on that assumption, the law should recognize this as de facto "actual knowledge." Therefore, judges should look beyond companies' basic claims or simple cop-outs like age screens. Instead, judges should look into what companies' data-governance capabilities actually are, including by considering the companies' own representations about those capabilities. They should insist upon consistency from tech companies and incentivize them to apply the same rigorous data-governance procedures in the contexts of both compliance and value-maximization.

The above, admittedly, is only a partial solution. Pushing for broader interpretations of existing law can only take us so far. Second, we propose to **adapt rules to disincentivize or disallow willful ignorance.** In addition to judges broadening their interpretations of existing law, policymakers must expand the scope of data-stewardship obligations to account for tech companies' increased data-governance capabilities. State-level privacy laws in recent years have taken a step forward in this regard.[158]

For example, Maryland's Age-Appropriate Design Code Act does not only trigger when a company has "actual

---

[156] Morris et al., *supra* note 24.
[157] Alicia Solow-Niederman, *Information Privacy and the Inference Economy*, 117 NW. U. L. REV. 357 (2022).
[158] Camille Altieri, *supra* note 152.

knowledge" that a child is using their online service, but also when they "know or should know" that is the case.[159] This creates room to consider each company's data-governance capabilities when determining potential liability. In doing so, it removes companies' incentives to feign ignorance.

Washington's My Health My Data Act, meanwhile, takes a different approach than HIPAA. There, the focus is not based on formalistic distinctions between providers, business associates, or nonbusiness associates.[160] Instead, health data itself is protected—regardless of whom it comes into contact with it.[161] Under a broader approach, for example, tech companies cannot eschew responsibility for the sensitive health data they receive simply because they are not technically a covered entity or business associate. Instead, the onus would fall on anyone in possession of the health data. Accordingly, this could incentivize those tracking patients to either intentionally filter out health data from their collection processes or, in the alternative, take steps to comply with the heightened legal protections.

In summary, in this Article, we demonstrate and analyze the hypocrisy of companies that make inconsistent representations of their data-governance capabilities when discussing compliance as compared to monetization. We advocate for transparency, through a combination of existing and emerging, technical and policy approaches.

Acknowledgements

This work was supported in part by the National Science Foundation (NSF) under grants numbers 1956393, 2138139, 2103439, and the UC Noyce Initiative and the Center for Information Technology Research in the Interest of Society (CITRIS) and the Banatao Institute. Olivia Figueira was also supported by the ARCS Danaher Foundation Fellowship and the UCI ICS Steckler Family Endowed Fellowship.

---

[159] MD. CODE ANN., COM. LAW § 14-4801 *et seq.* (2025).
[160] *See* WASH. REV. CODE. § 19.373 (2025).
[161] *Id.*