YOUR VOICE GAVE YOU AWAY: THE PRIVACY RISKS OF VOICE-INFERRED INFORMATION

EMMA RITTER†

ABSTRACT

Our voices can reveal intimate details about our lives. Yet, many privacy discussions have focused on the threats from speaker recognition and speech recognition. This Note argues that this focus overlooks another privacy risk: voice-inferred information. This term describes non-obvious information drawn from voice data through a combination of machine learning, artificial intelligence, data mining, and natural language processing. Companies have latched onto voice-inferred information. Early adopters have applied the technology in situations as varied as lending risk analysis and hiring. Consumers may balk at such strategies, but the current United States privacy regime leaves voice insights unprotected. By applying a notice and consent privacy model via sector-specific statutes, the hodgepodge of U.S. federal privacy laws allows voice-inferred information to slip through the regulatory cracks. This Note reviews the current legal landscape and identifies existing gaps. It then suggests two solutions that balance voice privacy with technological innovation: purpose-based consent and independent data review boards. The first bolsters voice protection within the traditional notice and consent framework, while the second imagines a new protective scheme. Together, these solutions complement each other to afford the human voice the protection it deserves.

† Duke University School of Law, J.D. expected 2022; Saint Olaf College, B.A. 2015. Thank you to all those who helped this Note become a reality. From my former co-workers who inspired the Note's driving question, to the *Duke Law Journal* editors and Professor Sarah Baker who offered thoughtful feedback, to my fiancé who provided endless patience during the writing process—this piece of scholarship owes its existence to the support of others.

INTRODUCTION

> *"When* I *use a word," Humpty Dumpty said in rather a scornful tone, "it means just what I choose it to mean—neither more nor less."*
>
> *"The question is," said Alice, "whether you* can *make words mean so many different things."*
>
> *"The question is," said Humpty Dumpty, "which is to be master—that's all."*
>
> *Lewis Carroll*[1]

Imagine you set aside an afternoon to conduct a number of important calls you had previously delayed. You call your health insurance company to ask about your deductible. Then, shuffling your papers, you call your bank to ask about a recent personal loan you took out. Finally, you open your laptop to complete a video interview for a potential new job. Clicking record, you answer a series of behavioral interview questions and upload the file to the company's HR portal. You close your laptop, to-do list complete.

Yet, these calls have a life that extends far beyond your afternoon task list. Each seemingly innocuous voice interaction has an outsized impact behind the scenes. Your healthcare company screened your voice for signs of Alzheimer's disease. Had the algorithm flagged you, your call agent would have offered resources for specialists in your area. Your bank analyzed your tone to assess the likelihood you would default on your loan. As a result, it placed you into a high-risk pool to closely monitor. Your HR recruiter analyzed your personality using your voice tone. Having determined you a poor culture fit, the company will likely move on to other candidates.

[1] LEWIS CARROLL, THROUGH THE LOOKING-GLASS 81–82 (1899).

This is no science fiction plot, however: these hypotheticals mirror reality.[2] Voice technology has exploded in recent years,[3] driven by advances in artificial intelligence.[4] These advances have made an impression on the private sector. One survey found that 69 percent of "IT decision-makers work at companies that currently invest in or plan to invest in voice technology within 3 years."[5] The *Harvard Business Review* recently urged businesses to invest in voice-first technology or risk "getting burned."[6] Facebook, Amazon, Microsoft, Google, and Apple have made sizeable investments in the voice space as they bet voice will become the next big platform.[7] These thought leaders recognize that voice technologies herald a radical change: by focusing on understanding consumer voice interactions, companies can create more "human" technologies that deliver better experiences.[8]

But privacy advocates warn the rise of voice technology has ushered in a new realm of privacy concerns.[9] These advocates often

[2] *See* Angela Chen, *Why Companies Want To Mine the Secrets of Your Voice*, VERGE (Mar. 14, 2019, 12:48 PM), https://www.theverge.com/2019/3/14/18264458 [https://perma.cc/X8D6-8GQ9] (reporting a large European bank used a voice start-up to categorize its debtors into risk pools); Matt Reynolds, *Health Insurer Calls Analysed for Signs of Disease in Your Voice*, NEWSCIENTIST (Feb. 6, 2017), https://www.newscientist.com/article/2120426 [https://perma.cc/K9B3-NKN9] (describing algorithms identifying early signs of Alzheimer's disease from phone calls to a health insurer); *Can Voice Analytics Help HR Find Better Candidates?*, SPEECH TECH. (Oct. 31, 2018), https://www.speechtechmag.com/Articles/ReadArticle.aspx?ArticleID=128280 [https://perma.cc/PDX4-R27J] (noting a company that offers HR departments an algorithm that uses speech parameters to determine candidate compatibility).

[3] *See Speech and Voice Recognition Market Worth $31.82 Billion by 2025*, GRAND VIEW RSCH. (Nov. 2018), https://www.grandviewresearch.com/press-release/global-voice-recognition-industry [https://perma.cc/EKJ7-9DTW] (forecasting voice technology growth at 17.2 percent).

[4] *See generally* Peng Lai "Perry" Li, *Natural Language Processing*, 1 GEO. L. TECH. REV. 98 (2016) (highlighting new technology used to increase accuracy of voice technology).

[5] APPDYNAMICS, THE FUTURE OF VOICE TECHNOLOGY IN THE ENTERPRISE 10 (emphasis omitted), https://cloud.kapostcontent.net/pub/7b04ec28-ccc9-4fd5-8eb2-6bad20c64128/the-future-of-voice-technology-in-the-enterprise [https://perma.cc/TF6S-3KQB].

[6] Bradley Metrock, *Your Company Needs a Strategy for Voice Technology*, HARV. BUS. REV. (Apr. 29, 2019), https://hbr.org/2019/04/your-company-needs-a-strategy-for-voice-technology-2 [https://perma.cc/W3N6-63UB].

[7] *How Big Tech Is Battling To Own the $49B Voice Market*, CB INSIGHTS (Feb. 13, 2019), https://www.cbinsights.com/research/facebook-amazon-microsoft-google-apple-voice [https://perma.cc/M3GW-67LC].

[8] For example, "[s]eniors often suffer from loneliness, isolation, and depression, and smart speakers have demonstrated effectiveness in counteracting this within nursing homes and senior living facilities." Metrock, *supra* note 6.

[9] *See, e.g.*, Dacia Green, Note, *Big Brother Is Listening to You: Digital Eavesdropping in the Advertising Industry*, 16 DUKE L. & TECH. REV. 352, 355–60 (2018) (discussing always-on

focus their criticism on speech recognition[10] and speaker recognition.[11] This focus makes some sense: voice assistants have proliferated, driven in part by the adoption of smart speakers in many homes.[12] But, while these technologies offer convenience, they also pose risks. Speech recognition technology can expose users to cyber threats and unwanted data sharing,[13] while speaker identification raises the specter of government overreach and identity theft. Each voice has a distinctive pattern as individual and identifiable as a fingerprint.[14] Unlike credit

devices and voice recording technology that allow companies to target digital advertising based on conversations).

[10] *See, e.g.*, Allison S. Bohm, Edward J. George, Bennett Cyphers & Shirley Lu, *Privacy and Liberty in an Always-On, Always-Listening World*, 19 COLUM. SCI. & TECH. L. REV. 1, 3–4 (discussing the intrusive effects of speech recognition in the home). "Speech recognition (also known as speech-to-text) is used by a large portion of people, with and without disabilities, in everything from mobile phones and GPS devices (to assist with hands-free calling and map directions) to dictation and controlling software applications." JONATHAN LAZAR, DANIEL GOLDSTEIN & ANNE TAYLOR, ENSURING DIGITAL ACCESSIBILITY THROUGH PROCESS AND POLICY 10 (2015).

[11] *See, e.g.*, Guangke Chen, Sen Chen, Lingling Fan, Xiaoning Du, Zhe Zhao, Fu Song & Yang Liu, *Who Is Real Bob? Adversarial Attacks on Speaker Recognition Systems*, 2021 IEEE SYMP. ON SEC. & PRIV. 694, 694 (noting the threat of voice identity theft for speaker recognition systems). Speaker recognition technology performs tasks such as identifying a speaker or detecting when two different recordings have the same speaker. Craig S. Greenberg, Lisa P. Mason, Seyed Omid Sadjadi & Douglas A. Reynolds, *Two Decades of Speaker Recognition Evaluation at the National Institute of Standards and Technology*, 60 COMPUT. SPEECH & LANGUAGE 1, 2 (2019).

[12] Sarah Perez, *Over a Quarter of US Adults Now Own a Smart Speaker, Typically an Amazon Echo*, TECHCRUNCH (Mar. 8, 2019, 12:29 PM), https://techcrunch.com/2019/03/08/over-a-quarter-of-u-s-adults-now-own-a-smart-speaker-typically-an-amazon-echo [https://perma.cc/F93U-XARD].

[13] *Smart Speaker Security—Tips To Make Sure Your Smart Speaker Is Secure*, KASPERSKY, https://usa.kaspersky.com/resource-center/threats/how-to-improve-your-smart-speaker-privacy [https://perma.cc/TEP3-CJFK]. Consumers share anecdotes of eavesdropping devices regularly; almost half of smart home device owners believe their devices record their conversations to better target ads. Sara Morrison, *Alexa Records You More Often Than You Think*, VOX (Feb. 21, 2020, 7:10 AM), https://www.vox.com/recode/2020/2/21/21032140 [https://perma.cc/ZG69-PRVT]; Alistair Charlton, *Half of US Adults Believe Smart Home Devices Record Conversations To Send Targeted Ads*, SALON (July 1, 2018, 5:29 PM), https://www.salon.com/2018/07/01/half-of-us-adults-believe-smart-home-devices-record-conversations-to-send-targeted-ads_partner [https://perma.cc/72TR-PJGX].

[14] Gary Audin, *Understand the Value of Voice Biometrics Basics*, NO JITTER (Sept. 4, 2020), https://www.nojitter.com/ai-speech-technologies/understand-value-voice-biometrics-basics [https://perma.cc/DZ7M-GWMW].

cards, compromised voiceprints cannot be replaced with a simple phone call to a financial institution.[15]

A more amorphous privacy threat also emerges from voice technology. Big data allows companies to "draw non-intuitive and unverifiable inferences and predictions about the behaviors, preferences, and private lives of individuals," such as "user preferences, sensitive attributes (e.g., race, gender, sexual orientation), . . . opinions (e.g., political stances), [and] behaviors (e.g., to serve advertisements)."[16] Unlike traditional data, this inferred data is created rather than collected. Seemingly meaningless data points, when combined, can create sensitive insights.[17] Companies value this data highly. Speaker recognition can increase call center security and efficiency,[18] and speech recognition enables integrated voice response menus that lower call center costs.[19] But voice-inferred information—what this Note refers to as "voice insights"—outstrips both.

Voice insights come with innovative promise and privacy threats. Like any inferred information, voice insights will allow companies to achieve high opportunity use cases.[20] For example, a company could

[15] Paul Mee & Gokhanedge Ozturk, *Prepare To Protect Your Customers' Voices*, MIT SLOAN MGMT. REV. (May 5, 2020), https://sloanreview.mit.edu/article/prepare-to-protect-your-customers-voices [https://perma.cc/422U-NRVU].

[16] Sandra Wachter & Brent Mittelstadt, *A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI*, 2019 COLUM. BUS. L. REV. 494, 497; *see also* Jacob Leon Kröger, Otto Hans-Martin Lutz & Philip Raschke, *Privacy Implications of Voice and Speech Analysis—Information Disclosure by Inference*, *in* PRIVACY AND DATA MANAGEMENT 242, 242 (Kai Rannenberg ed., 2020) (cataloging the wide range of sensitive information that researchers can ascertain from the human voice).

[17] Sheri B. Pan, *Get To Know Me: Protecting Privacy and Autonomy Under Big Data's Penetrating Gaze*, 30 HARV. J.L. & TECH. 239, 248 (2016) (discussing how "liking" MAC Cosmetics on Facebook can predict sexual orientation, a highly sensitive trait).

[18] Audin, *supra* note 14.

[19] *See* Severine Griziaux, *The Seven Benefits of an IVR System*, TWILIO, https://www.twilio.com/learn/voice-and-video/the-seven-benefits-of-an-ivr-system [https://perma.cc/8RCZ-MU6F] ("[C]ompared to live chat at $5 per contact, or telephone-based customer service that ranges from $6 to over $12 per contact, an [integrated voice response, or "IVR,"] can cost less than $1 per contact. . . ."). Readers are likely familiar with IVRs, the (often aggravating) recorded menus that listen to your voice to direct the route of your call.

[20] Common among programmers and product teams, the term "use case" denotes a specific way a user engages with a system to achieve a specific goal. *Use Case*, TECHOPEDIA, https://www.techopedia.com/definition/25813/use-case [https://perma.cc/GA25-NA6R]. Voice insights represent a "high opportunity use case" because of the high value to companies they represent: the global speech and voice recognition market "was valued at USD 6.9 Billion in 2018," and will "reach a value of USD 28.3 Billion by the end of 2026." *Speech and Voice Recognition Market To Be Worth USD 28.3 Billion by 2026, Rising at a CAGR of 19.8%*, FORTUNE BUS. INSIGHTS (Apr.

use data, including voice data, to optimize pricing in real time or to create radically personalized products.[21] At the same time, voice insights may open the door to inaccurate and discriminatory personalization.[22] Of course, companies do not, and will not, use voice insights for nefarious objectives only.[23] Still, voice insights pose enough risk to warrant closer scrutiny.

The United States' current data privacy schema may not prove up to the challenge of regulating voice insights: the United States lacks a general federal data privacy law.[24] Unlike the European Union ("EU"), which recently enacted the General Data Protection Regulation ("GDPR"),[25] the United States employs a patchwork of

22, 2021, 7:26 AM), https://www.globenewswire.com/en/news-release/2021/04/22/2214930/0/en/Speech-and-Voice-Recognition-Market-to-be-Worth-USD-28-3-Billion-by-2026-Rising-at-a-CAGR-of-19-8.html [https://perma.cc/7JNC-SCGJ].

[21] *See generally* NICOLAUS HENKE, JACQUES BUGHIN, MICHAEL CHUI, JAMES MANYIKA, TAMIM SALEH, BILL WISEMAN & GURU SETHUPATHY, MCKINSEY GLOB. INST., THE AGE OF ANALYTICS: COMPETING IN A DATA-DRIVEN WORLD (2016) ("Hyperscale digital platforms can match buyers and sellers in real time, transforming inefficient markets. Granular data can be used to personalize products and services . . . .").

[22] Pan, *supra* note 17, at 250–52.

[23] For example, at least one voice technology company has emphasized the potential for its product to help eradicate unconscious bias in call center transactions. *See Reducing Bias in Customer Engagement*, COGITO, https://cogitocorp.com/resources/on-demand-webinar-reducing-bias/enjoy-the-webinar (last visited Oct. 1, 2021) (highlighting its use of real-time "nudges" to help call center agents form better connections with customers).

[24] Shawn Marie Boyne, *Data Protection in the United States*, 66 AM. J. COMP. L. 299, 299 (2018).

[25] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1–88 [hereinafter GDPR]; *see* Ben Wolford, *What Is GDPR, the EU's New Data Protection Law?*, GDPR.EU, https://gdpr.eu/what-is-gdpr [https://perma.cc/SJ5R-BR4W] (providing an overview of GDPR). The GDPR extends beyond the EU and encompasses the entire "European Economic Area (EEA), which includes all EU countries plus Iceland, Liechtenstein and Norway. When personal data is transferred outside the EEA, the protections offered by the GDPR should travel with the data. This means that to export data abroad, companies must ensure that certain safeguards are in place." EUR. COMM'N, THE GDPR: NEW OPPORTUNITIES, NEW OBLIGATIONS 15 (2018), https://ec.europa.eu/info/sites/default/files/data-protection-factsheet-sme-obligations_en.pdf [https://perma.cc/NVJ6-7KR4].

sector-based federal laws[26] and state privacy laws to protect data.[27] U.S. privacy laws apply to specific situations involving "healthcare, education, communications, and financial services or, in the case of online data collection, to children."[28] Where data collection or use avoids these narrow confines, no law is implicated.[29] The GDPR, in contrast, applies whenever personal data processing relates to the offering of goods or services to or the monitoring of individual behavior of European citizens or residents.[30]

The Fair Information Practice Principles ("FIPPs") provide a framework for assessing privacy protections. Originally introduced in a 1970s report by the U.S. Department of Health, Education, and Welfare,[31] these principles have spread beyond the U.S. border. In 1980, the Organisation for Economic Co-operation and Development ("OECD") "revised the principles in an internationally influential document that continues to serve as the bedrock foundation for privacy

[26] *See* Boyne, *supra* note 24 ("Privacy protection guarantees are sector-specific and are located in a myriad of legislative instruments and case law."). "Sector" refers to specific industries or areas of economic activity; for example, a loan company would fall into the financial sector. *Sector*, CAMBRIDGE DICTIONARY, https://dictionary.cambridge.org/us/dictionary/english/sector [https://perma.cc/589V-5296]. *See generally infra* Part II.

[27] *See Security Breach Notification Laws*, NAT'L CONF. OF STATE LEGISLATURES, https://www.ncsl.org/research/telecommunications-and-information-technology/security-breach-notification-laws [https://perma.cc/M9WZ-4BAH] (last updated Apr. 15, 2021) (reporting that "[a]ll 50 states, the District of Columbia, Guam, Puerto Rico and the [U.S.] Virgin Islands" have security breach laws for personally identifiable information).

[28] Boyne, *supra* note 24 (quoting N. Terry, *Existential Challenges for Health Care Data Protection in the United States*, 3 ETHICS, MED. & PUB. HEALTH 19, 21 (2017)).

[29] *Id.*

[30] GDPR, *supra* note 25, at 33. Examples of offering goods or services to EEA citizens include "creat[ing] ads in German or includ[ing] pricing in euros on its website"; examples of monitoring behavior include "us[ing] web tools that allow you to track cookies or the IP addresses of people who visit your website from EU countries." Ben Wolford, *Does the GDPR Apply to Companies Outside of the EU?*, GDPR.EU, https://gdpr.eu/companies-outside-of-europe [https://perma.cc/53XF-QMKU]. An omnipresent feature of the internet, "[c]ookies are small files that websites send to your device that the sites then use to monitor you and remember certain information about you — like what's in your shopping cart on an e-commerce site, or your login information." Emily Stewart, *Why Every Website Wants You To Accept Its Cookies*, VOX (Dec. 10, 2019, 8:00 AM), https://www.vox.com/recode/2019/12/10/18656519/what-are-cookies-website-tracking-gdpr-privacy [https://perma.cc/BAU7-ULYL].

[31] WOODROW HARTZOG, PRIVACY'S BLUEPRINT 59 (2018). *See generally* U.S. DEP'T OF HEALTH, EDUC. & WELFARE, RECORDS, COMPUTERS, AND THE RIGHTS OF CITIZENS: REPORT OF THE SECRETARY'S ADVISORY COMMITTEE ON AUTOMATED PERSONAL DATA SYSTEMS (1973), https://www.justice.gov/opcl/docs/rec-com-rights.pdf [https://perma.cc/294D-AJAK] (presenting the original 1970s report).

regulatory schemes and public policy"[32] around the world.[33] While not legally binding, the FIPPs detail best practices for data protection laws.[34] Today, the FIPPs shape data protection regimes "in Australia, Canada, the European Union, and many Asian countries," as well as in the United States.[35]

The FIPPs outline eight basic principles: collection limitation, data quality, purpose specification, use limitation, security safeguards, openness, individual participation, and accountability.[36] Each principle protects personal data in a different way. For example, the purpose specification principle calls for organizations to specify "[t]he purposes for which personal data are collected . . . not later than at the time of data collection."[37] Subsequent use of the data must be limited to the same or compatible purposes, and the organization must specify the change of purpose on each occasion.[38] Countries can combine these principles in different ways and to varying degrees in order to create a data protection regime.

The United States employs use and collection limitations focused on specific economic sectors.[39] The use limitation principle requires that "[p]ersonal data should not be disclosed, made available or otherwise used . . . except: a) *with the consent of the data subject*; or b) by the authority of law."[40] Collection limitation, meanwhile, advocates "limits to the collection of personal data" and prescribes that "any such data should be obtained by lawful and fair means and, where

[32] HARTZOG, *supra* note 31, at 59. The OECD updated the original guidelines from 1980 in 2013. OECD, THE OECD PRIVACY FRAMEWORK 3 (2013) [hereinafter THE OECD PRIVACY FRAMEWORK], http://www.oecd.org/sti/ieconomy/oecd_privacy_framework.pdf [https://perma.cc/DL85-3BQ6].

[33] Frederik Zuiderveen Borgesius, Jonathan Gray & Mireille van Eechoud, *Open Data, Privacy, and Fair Information Principles: Towards a Balancing Framework*, 30 BERKELEY TECH. L.J. 2073, 2104 (2015).

[34] *Id.* at 2103.

[35] HARTZOG, *supra* note 31, at 60.

[36] THE OECD PRIVACY FRAMEWORK, *supra* note 32, at 14–15 (outlining the basic principles with national application).

[37] *Id.* at 14.

[38] *Id.*

[39] *See* Boyne, *supra* note 24, at 299 ("[L]egislation at the federal level primarily protects data within sector-specific contexts.").

[40] THE OECD PRIVACY FRAMEWORK, *supra* note 32, at 14 (emphasis added).

appropriate, *with the knowledge or consent of the data subject*."[41] Both principles thus focus on consumer notice and consent.[42]

This Note contends that current data privacy laws in the United States do not adequately address voice-inferred information. As calls mount for a federal data protection law,[43] commentators have begun to sketch the contours of this future comprehensive legislation.[44] This Note contributes to the discussion by focusing on the privacy threat posed by voice insights. It evaluates the current U.S. model of notice and consent in light of this threat, ultimately advocating for the adoption of purpose specification principles in any comprehensive federal privacy law.[45]

To support this conclusion, this Note begins with an overview of emerging technologies relevant to the voice privacy regulation discussion. Part I walks through the high-level mechanics of big data, predictive analytics, and voice technology. Part II surveys the current regulatory landscape within the United States, noting each law's ability to protect voice insights. Part III argues for the adoption of purpose-based limitations to protect voice data privacy. It offers two policy solutions by which a federal law could implement purpose specification requirements: purpose-based consent and independent data review boards.

[41] *Id.* (emphasis added).

[42] These two principles form the bedrock of U.S. privacy policy. Woodrow Hartzog & Neil Richards, *Privacy's Constitutional Moment and the Limits of Data Protection*, 61 B.C. L. REV. 1687, 1700 (2020).

[43] Commentators view "U.S. privacy law [as] in the midst of a . . . period of unusual public engagement likely to result in a significant and durable settlement of the issues." *Id.* at 1694; *see, e.g.*, Jessica Rich, *After 20 Years of Debate, It's Time for Congress To Finally Pass a Baseline Privacy Law*, BROOKINGS (Jan. 14, 2021), https://www.brookings.edu/blog/techtank/2021/01/14/after-20-years-of-debate-its-time-for-congress-to-finally-pass-a-baseline-privacy-law [https://perma.cc/5AEY-Q7DM] (advocating for a comprehensive law to protect the nation's privacy and cybersecurity).

[44] *See, e.g.*, Jay P. Kesan, Carol M. Hayes & Masooda N. Bashir, *A Comprehensive Empirical Study of Data Privacy, Trust, and Consumer Autonomy*, 91 IND. L.J. 267, 346–49 (2016) (outlining one potential strategy coupling government oversight with the creation of third-party profile repository agencies).

[45] Here, comprehensive simply means these statutes do not limit themselves to a single sector.

I. BIG DATA AND VOICE TECHNOLOGY

Data privacy protections must match the pace of technological advances to remain effective. This Part provides high-level background information on the technology driving voice insights: big data and predictive analytics. Part I.A discusses the rise of big data, while Part I.B explains how analytics have developed to parse and process enormous amounts of data. Part I.C then details how such analytics apply in the context of voice data. Together, these sections elucidate the technology any data privacy regulation must address.

A. Quantifying Big Data: How Big is Big?

Data permeates our lives. In 2019, global data reached 45 zettabytes of information.[46] A zettabyte measures storage capacity; it represents 1 sextillion bytes (1,000,000,000,000,000,000,000 bytes).[47] Historically, data existed as an output from scientific studies aimed at collecting and analyzing trends, such as "predict[ing] the movements of the sun and stars and determin[ing] population-wide rates of crime, marriage, and suicide."[48] Today, data enjoys a far wider array of origins. Sources "include information systems, digitalization, sensors, surveillance and tracking systems, the [Internet of Things],[49] mobile devices and applications, social services and network platforms, and wearable . . . devices and services."[50] Data sources will continue to increase.[51] The International Data Corporation has predicted that by

[46] DAVID REINSEL, JOHN GANTZ & JOHN RYDNING, DATA AGE 2025: THE DIGITIZATION OF THE WORLD FROM EDGE TO CORE 6 (2020).

[47] Thomas Barnett, Jr., *The Zettabyte Era Officially Begins (How Much Is That?)*, CISCO BLOGS (Sept. 9, 2016), https://blogs.cisco.com/sp/the-zettabyte-era-officially-begins-how-much-is-that [https://perma.cc/8ZP9-RP3P].

[48] EXEC. OFF. OF THE PRESIDENT, BIG DATA: SEIZING OPPORTUNITIES, PRESERVING VALUES 1 (2014), https://obamawhitehouse.archives.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf [https://perma.cc/79GG-32FB].

[49] The Internet of Things is a term that encompasses "objects that 'talk' to each other." Matt Burgess, *What Is the Internet of Things? WIRED Explains*, WIRED (Feb. 16, 2018, 12:40 PM), https://www.wired.co.uk/article/internet-of-things-what-is-explained-iot [https://perma.cc/HT6E-7L3V]. Can you control your toaster settings from your phone? Does your dog's collar send you health notifications? If so, you have used the Internet of Things.

[50] *See* Longbing Cao, *Data Science: A Comprehensive Overview*, 50 ACM COMPUTING SURVS. 43:1, 43:9 (2017), https://dl.acm.org/doi/pdf/10.1145/3076253 [https://perma.cc/J7Z6-W5XS] (describing big data's increasing ability to quantify data from any source).

[51] *Cf.* Luke Fitzpatrick, *The "Rise of Alternative Data:" So, What the Heck Is It?*, CPO MAG. (Feb. 7, 2020), https://www.cpomagazine.com/data-privacy/the-rise-of-alternative-data-so-what-

2025 the global datasphere will reach 175 zettabytes, more than triple the amount today.[52]

Data itself generates data, known as metadata. Metadata explains "the layout and meaning of the data"[53] by "describ[ing] properties of the data such as the time the data were created, the device on which they were created, or the destination of a message."[54] Think about a photo stored as a computer file. The photo's metadata might include information like the size of the photo, when the photo was taken, and who took the photo.[55] Similarly, an email might include metadata "about the sending and destination addresses" or even "the routing of the path between them."[56] Metadata allows for better identification, use, and re-use of data.[57] It operationalizes data sources that would otherwise prove too unwieldy, helping organizations grapple with enormous quantities of data.[58]

The "big data" moniker reflects this rapid expansion of data. Definitions of big data abound,[59] but every definition references at least one of the following factors: size, complexity, and technologies.[60] Combining these factors, big data becomes "the storage and analysis of

the-heck-is-it [https://perma.cc/5UWR-HEQY] (describing the increase in business use of new data sources).

[52] REINSEL ET AL., *supra* note 46, at 6.

[53] EXEC. OFF. OF THE PRESIDENT, BIG DATA AND PRIVACY: A TECHNOLOGICAL PERSPECTIVE 19 (2014) [hereinafter BIG DATA AND PRIVACY], https://bigdatawg.nist.gov/pdf/pcast_big_data_and_privacy_-_may_2014.pdf [https://perma.cc/6FCH-Z8PQ].

[54] *Id.* at xi. Although metadata is data about data, it can still pose a privacy threat because it includes personal information such as "account numbers, login names, and passwords." *Id.* at 19.

[55] *Working with Metadata in Images*, ORACLE HELP CTR., https://docs.oracle.com/cd/B19306_01/appdev.102/b14302/ch_metadata.htm [https://perma.cc/FJ66-69CC].

[56] BIG DATA AND PRIVACY, *supra* note 53, at 19 n.53.

[57] *See, e.g.*, *Metadata and Its Importance in a Data Driven World*, VILL. UNIV., https://www.villanovau.com/resources/bi/metadata-importance-in-data-driven-world [https://perma.cc/6GU9-C2JF] (last updated Oct. 24, 2019) (comparing metadata to "effective cataloging" that allows for effective organization of data and increases its interoperability).

[58] *Id.*

[59] *See* Jonathan Stuart Ward & Adam Barker, Undefined by Data: A Survey of Big Data Definitions 1 (Sept. 20, 2013) (unpublished manuscript), https://arxiv.org/pdf/1309.5821v1.pdf [https://perma.cc/66X2-TUVS] (noting that big data's use by academia, business, media, and other stakeholders has created "diverse and often contradictory definitions" of the term).

[60] *See id.* at 2 (describing commonalities in definitions surveyed). Here, "technologies" refers to "the tools and techniques . . . used to process a sizable or complex dataset." *Id.*

large and or complex data sets using a series of techniques" adapted specifically to large quantities of data.[61]

B. Putting Big Data to Use: The Spectrum of Data Analytics

Big data is only valuable when it can be analyzed.[62] Some technologies, like artificial intelligence ("AI") and machine learning,[63] have developed specifically to allow analysis of big data.[64] To fully understand these technologies, however, it is helpful to understand the different types of data analysis and the trajectory of the data science field.

Data scientists commonly cite four "different types of analytics": "descriptive, diagnostic, predictive, and prescriptive."[65] Descriptive and diagnostic analytics deal with the past.[66] Specifically, descriptive analytics performs statistical analysis on data; it asks *what* happened.[67] Diagnostic analytics, on the other hand, performs root cause analysis; it looks at *how* and *why* something happened.[68] Typical outputs from descriptive and diagnostic analytics include graphs, charts, and

[61] *Id.* Several articles cite this definition when referring to big data. *See, e.g.*, Max N. Helveston, *Consumer Protection in the Age of Big Data*, 93 WASH. U. L. REV. 859, 868 (2016) (adopting the definition proposed by Ward and Barker).

[62] *See* Charles Arthur, *Tech Giants May Be Huge, but Nothing Matches Big Data*, GUARDIAN (Aug. 23, 2013, 3:21 PM), https://www.theguardian.com/technology/2013/aug/23/tech-giants-data [https://perma.cc/95U6-J9EM] ("Data is just like crude [oil]. It's valuable, but if unrefined it cannot really be used. It has to be changed into gas, plastic, chemicals, etc to create a valuable entity that drives profitable activity; so must data be broken down, analysed for it to have value.").

[63] *See infra* notes 80–91 and accompanying text.

[64] "[M]astering data is insurmountable without AI." *Big Data and Artificial Intelligence: How They Work Together*, MARYVILLE UNIV., https://online.maryville.edu/blog/big-data-is-too-big-without-ai [https://perma.cc/4BRA-2JS6]. And "machine learning is a . . . subset of AI." Wayne Thompson, Hui Li & Alison Bolen, *Artificial Intelligence, Machine Learning, Deep Learning and Beyond*, SAS, https://www.sas.com/en_us/insights/articles/big-data/artificial-intelligence-machine-learning-deep-learning-and-beyond.html [https://perma.cc/LE9P-PUVP].

[65] Brian Brinkmann, *Comparing Descriptive, Predictive, Prescriptive, and Diagnostic Analytics*, LOGI ANALYTICS, https://www.logianalytics.com/predictive-analytics/comparing-descriptive-predictive-prescriptive-and-diagnostic-analytics [https://perma.cc/F78X-WVUJ] (last updated Feb. 18, 2021).

[66] *See id.* (describing the insights diagnostic and descriptive analytics provide about past data).

[67] *Id.*

[68] *Id.*

dashboards with the ability to drill through[69] to gain more information about a finding.[70]

For example, consider a healthcare setting. Descriptive analytics can identify that "an unusually high number of people [were] admitted to the emergency room in a short period of time . . . [and provide] corresponding statistics (date of occurrence, volume, patient details, etc.)."[71] Diagnostic analytics can "determine that all of the patients' symptoms—high fever, dry cough, and fatigue—point to the same infectious agent."[72] While descriptive and diagnostic analytics can help companies, they have drawbacks.[73] Both categories keep businesses in a reactive mode; they analyze only what companies know they need to address.[74]

In contrast, predictive and prescriptive analytics look to the future. Predictive analytics uses past data to predict future events; it asks *what will* happen.[75] Prescriptive analytics "suggests various courses of action and outlines what the potential implications would be for each"; it asks *what the next best action* is.[76] Typical outputs for predictive and prescriptive analytics include "predictive modeling, optimization, . . . and actionable knowledge delivery."[77] Returning to the healthcare example, predictive analytics allow the hospital to "forecast a surge in patients admitted to the ER in the next several weeks," while prescriptive analytics "may suggest that you increase the number of staff on hand to adequately treat the influx of patients."[78] Descriptive and diagnostic analytics still play a role in this scenario. But an

[69] Drill through data reports allow businesses to navigate between different views of connected data. *Drill Through Access*, IBM: COGNOS ANALYTICS, https://www.ibm.com/docs/en/cognos-analytics/11.1.0?topic=reporting-drill-through-access [https://perma.cc/3KKE-JRQM]. For example, a user might click on a specific point in time on a line graph to view another graph tracking which products constituted the bulk of that month's sales, among other possibilities. *Id.*

[70] Brinkmann, *supra* note 65.

[71] *Id.*

[72] *Id.*

[73] *See* Cao, *supra* note 50, at 43:20 (describing the limitations of explicit analytics that deal only with known unknowns, and attributing the shift to deep analytics to these limitations).

[74] *Id.* at 43:17–18.

[75] Brinkmann, *supra* note 65.

[76] *Id.*

[77] *See* Cao, *supra* note 50, at 43:20 (describing typical approaches to deep analytics, which encompasses predictive and prescriptive analytics). Optimization identifies the best option among a variety of approaches, while actionable knowledge delivery recommends specific actions to take for business decision-making and operations. *Id.* at 43:19.

[78] Brinkmann, *supra* note 65.

organization can automate these lower-level analytics, enabling its analysts to focus on the higher-value predictive and prescriptive analytics.[79]

Big data challenges the manual nature of traditional data analysis, thereby impelling the use of technologies like machine learning.[80] Big data gathers large amounts of high-dimensional[81] data "from multiple sources at different time points using different technologies."[82] Traditional statistics cannot handle the analytical and interpretive challenges these attributes create.[83] But emerging technologies such as artificial intelligence ("AI") are especially equipped to handle high "volumes, velocities and variety of data."[84] Machine learning, a subset of AI, "automates analytical model building" that can crunch the astounding numbers traditional statistics cannot.[85] Both technologies enable organizations to work with increasingly large datasets.

Machine learning also facilitates the discovery of inferred information, often through data mining. Data mining draws on machine learning to "discover[] patterns in large data sets."[86] It creates inferences. These inferred patterns can involve both past and future data, allowing data mining to play both a descriptive and predictive

[79] *See* Cao, *supra* note 50, at 43:20–21 (explaining the stages of the "paradigm shift" in analytics).

[80] Usama Fayyad, Gregory Piatetsky-Shapiro & Padhraic Smyth, *From Data Mining to Knowledge Discovery in Databases*, A.I. MAG., Fall 1996, at 37, 37–38.

[81] High-dimensional data has more than ten attributes. JIAWEI HAN, MICHELINE KAMBER & JIAN PEI, DATA MINING: CONCEPTS AND TECHNIQUES 508 (3d ed. 2012). Consider a store that carries tens of thousands of products; a customer's purchase profile would correspondingly include tens of thousands of dimensions to track which products the customer has purchased. *Id.* at 509.

[82] Jianqing Fan, Fang Han & Han Liu, *Challenges of Big Data Analysis*, 1 NAT'L SCI. REV. 293, 294 (2014).

[83] *See generally id.* (addressing the technical limitations of traditional statistics when faced with large sample sizes, high heterogeneity, spurious correlations, and other analytical challenges).

[84] Daniel E. O'Leary, *Artificial Intelligence and Big Data*, 28 IEEE INTELLIGENT SYS. 96, 97 (2013).

[85] Thompson et al., *supra* note 64.

[86] BIG DATA AND PRIVACY, *supra* note 53, at 24. Data mining borrows from fields other than machine learning as well. Liane Colonna, *A Taxonomy and Classification of Data Mining*, 16 SMU SCI. & TECH. L. REV. 309, 314 (2013) (noting data mining's reliance on "statistics, visualization, [and] pattern recognition"). Professors Igor Kononenko and Matjaz Kukar offer a helpful distinction between machine learning and data mining: "[w]hile machine learning focuses more on development of data modelling techniques, data mining is more application-oriented." IGOR KONONENKO & MATJAZ KUKAR, MACHINE LEARNING AND DATA MINING 34 (2007).

analytical role.[87] For example, data mining may identify that "people who live under high-voltage power lines have higher morbidity."[88] Of course, statistics counsels that correlation does not equal causation. While the data may indicate the health threat of power lines, it may also simply reflect a lack of healthcare access among the individuals with lower socioeconomic status living there.[89] Inferred information can empower organizations.[90] Given its "non-intuitive," "non-verifiable" nature, however, inferred information can also create new privacy hurdles where it introduces biases or infers sensitive information.[91]

C. Applying Analytics to Voice: An Overview of Voice Technologies

Voice technology builds on existing big data and analytics techniques to address difficulties specific to voice data. Consider a survey that requires participants to respond to three different questions. One question asks participants to select one of three provided options; one prompts participants to enter their own thoughts into a text box; one allows participants to upload a short audio recording to answer the question. With each data type, the processing difficulty increases. The first comes ready to analyze: a data analyst can quickly put the data into a graph or other visualization to see participant distribution across the three options.

Free-form text presents more difficulty. Participant answers will vary, requiring the analyst to process the data in some way before analysis can begin.[92] With big data, this processing becomes much more

[87] *See* Fayyad et al., *supra* note 80, at 44 ("The two high-level primary goals of data mining in practice tend to be prediction and description.").

[88] BIG DATA AND PRIVACY, *supra* note 53, at 25.

[89] *Id.*

[90] *See* Yeslam Al-Saggaf, *The Use of Data Mining by Private Health Insurance Companies and Customers' Privacy: An Ethical Analysis*, 24 CAMBRIDGE Q. HEALTHCARE ETHICS 281, 282 (2015) (describing how data mining can assist in identifying fraud and underdiagnosed patients in the healthcare setting).

[91] Wachter & Mittelstadt, *supra* note 16, at 497.

[92] *See* Daniel Martin, *Tapping the Value of Unstructured Data: Challenges and Tools To Help Navigate*, DATAVERSITY (Feb. 24, 2021), https://www.dataversity.net/tapping-the-value-of-unstructured-data-challenges-and-tools-to-help-navigate [https://perma.cc/N93M-GU7V] ("There are multiple challenges faced while working with unstructured data, namely . . . [m]ore processing is required.").

time- and labor-intensive.[93] Text analytics automates this process by enabling a computer program to "uncover[] insights such as sentiment analysis, entities, relations and key phrases in unstructured text."[94] This analysis relies on natural language processing ("NLP") and machine learning.[95] NLP allows "a computer to analyze what a user *said* . . . and process what the user *meant*,"[96] extracting the data points data analysts need to conduct their inquiry.

Like free-form text, voice data requires processing before analysis can take place.[97] But voice data adds further complexity: before an NLP system can analyze the meaning of a word or sentence, it needs to recognize the word or phrase in the first place.[98] This means taking into account speaker pronunciation, inflection, and timing.[99] Speech-to-text programs transcribe audio data to text for ease of use,[100] but reducing voice data to a transcription removes valuable information. Vocal features like "pitch, loudness, and the presence and duration of speech pauses . . . can reveal both state- and trait-level information about a speaker."[101] Perhaps the third question in the hypothetical survey asked about user satisfaction. Speech-to-text might categorize two users who respond, "I just *love* your service," as "Very Satisfied."

[93] For example, a team might task an intern to read through each response, create categories, and match each response to a category. The intern may only need a few minutes if the survey gained a handful of responses, but tagging a survey with thousands of participants could quickly snowball into a multi-hour process—all before analysis can start. *Cf. id.* (describing the challenges of processing unstructured data like free-form text).

[94] *Text Analytics*, MICROSOFT AZURE, https://azure.microsoft.com/en-us/services/cognitive-services/text-analytics [https://perma.cc/7UB5-XQR5].

[95] Kevin D. Ashley, *Automatically Extracting Meaning from Legal Texts: Opportunities and Challenges*, 35 GA. ST. U. L. REV. 1117, 1117 (2019) (describing legal text analytics as applying these computational techniques).

[96] Li, *supra* note 4, at 98.

[97] *Id.* at 99–100.

[98] *Id.* at 99–101.

[99] *Id.* (explaining the different abstraction levels an NLP system must process to recognize, parse, and understand data, including the particularities of human speech).

[100] *See, e.g.*, *Speech to Text*, MICROSOFT AZURE, https://azure.microsoft.com/en-us/services/cognitive-services/speech-to-text [https://perma.cc/UDR5-P258] (promoting speech-to-text as a way to "[g]et more value from spoken audio by enabling search or analytics on transcribed text" in order to "[m]ake spoken audio actionable").

[101] Christian Hildebrand, Fotis Efthymiou, Francesc Busquet, William H. Hampton, Donna L. Hoffman & Thomas P. Novak, *Voice Analytics in Business Research: Conceptual Foundations, Acoustic Feature Extraction, and Applications*, 121 J. BUS. RSCH. 364, 364 (2020).

Voice analytics, however, could catch that one user spoke sincerely, while the other's voice dripped with sarcasm.[102]

The potential—and the peril—of voice analytics is that it goes beyond what a human listener can catch; our voice gives away far more insights than we realize. We might expect that software can detect a caller's "heightened emotional state, either positive or negative," the same way a human call agent might.[103] Speaker identification, too, mirrors our human ability to recognize individual voices.[104] But speech patterns can also reveal physical[105] and mental illness.[106] Voice data conveys clues to "a speaker's biometric identity, personality, physical traits, geographical origin, emotions, level of intoxication and sleepiness, age, gender, and health condition," along with socioeconomic status in certain speech patterns.[107]

And yet, people cannot avoid using their voices. The human voice enables communication in a wide range of situations, from friendly banter to customer service. Speaking to another person remains

[102] *Cf. id.* at 366, 366 fig.2 (displaying the distinguishable difference between a normal "Hello" and one excitedly said to greet a close friend).

[103] Tom Simonite, *This Call May Be Monitored for Tone and Emotion*, WIRED (Mar. 19, 2018, 7:00 AM), https://www.wired.com/story/this-call-may-be-monitored-for-tone-and-emotion [https://perma.cc/G4MS-NZF7]. Call agents at Metropolitan Life Insurance Company receive "nudge[s]" from a machine-learning powered "empathy adviser" that, for example, may suggest to "[c]alm down" when dealing with a worked-up caller or may offer "soothing talking points." *Id.*

[104] *See* Hildebrand et al., *supra* note 101, at 372 (describing how banking, law enforcement, and other industries "are beginning to use voice samples as a consumer identification tool").

[105] *See, e.g.*, Resul Das, *A Comparison of Multiple Classification Methods for Diagnosis of Parkinson Disease*, 37 EXPERT SYS. WITH APPLICATIONS 1568, 1572 (2010) (finding a Neural Networks model identified Parkinson's disease with a 92.9 percent success rate using biomedical voice data).

[106] *See* Charles R. Marmar, Adam D. Brown, Meng Qian, Eugene Laska, Carole Siegel, Meng Li, Duna Abu-Amara, Andreas Tsiartas, Colleen Richey, Jennifer Smith, Bruce Knoth & Dimitra Vergyri, *Speech-Based Markers for Posttraumatic Stress Disorder in US Veterans*, 36 DEPRESSION & ANXIETY 607, 607 (2019) (demonstrating that "a speech-based algorithm can objectively differentiate PTSD cases from controls" through markers indicating "slower, more monotonous speech, less change in tonality, and less activation"); Skyler Place, Danielle Blanch-Hartigan, Channah Rubin, Cristina Gorrostieta, Caroline Mead, John Kane, Brian P. Marx, Joshua Feast, Thilo Deckersbach, Alex Pentland, Andrew Nierenberg & Ali Azarbayejani, *Behavioral Indicators on a Mobile Sensing Platform Predict Clinically Validated Psychiatric Symptoms of Mood and Anxiety Disorders*, 19 J. MED. INTERNET RSCH. *1, *6 (2017) (predicting "clinician-assessed symptoms of depressed mood," including fatigue and social disconnectedness).

[107] Kröger et al., *supra* note 16, at 242.

customers' preferred way to answer a question.[108] As voice data collection spreads, consumers will have to choose between protecting their voice data and receiving customer services from companies. Voice surveillance has entered the home as well. Always-listening devices perch atop many kitchen counters,[109] but even household appliances like refrigerators and TVs have begun to record and monitor conversations.[110] As voice technology proliferates, privacy laws must instead protect the data encapsulated in the human voice.

II. Regulatory Landscape

The rise of inferred information threatens the efficacy of the United States' sector-specific collection and use limitations.[111] With the rise of big data, data no longer fits into neat sectoral categories. Even "innocuous data about a person" can enable "inferences of a sensitive nature."[112] Voice insights pose similar problems. Speech patterns themselves do not implicate any particular sector, even if the content of a conversation might. This Part discusses the current patchwork of U.S. data privacy laws, focusing specifically on the FIPPs underlying each. Part II.A provides an overview of major sectoral-based statutes, which rely on use limitations. Part II.B looks at broader laws, which do not limit themselves to a specific sector.[113] The strengths and

[108] Gregg Johnson, *Your Customers Still Want To Talk to a Human Being*, HARV. BUS. REV. (July 26, 2017), https://hbr.org/2017/07/your-customers-still-want-to-talk-to-a-human-being [https://perma.cc/TKA4-SEWN] (reporting that most consumers still prefer to call a business when considering a high-value purchase or grappling with a question).

[109] *See* Perez, *supra* note 12 and accompanying text.

[110] Indeed, Samsung issued its Smart TV with a warning "that if [a person's] spoken words include personal or other sensitive information, that information will be among the data captured and transmitted to a third party through [their] use of Voice Recognition." Chris Matyszczyk, *Samsung's Warning: Our Smart TVs Record Your Living Room Chatter*, CNET (Feb. 8, 2015, 2:10 PM), https://www.cnet.com/news/samsungs-warning-our-smart-tvs-record-your-living-room-chatter [https://perma.cc/K5Z2-NAA2].

[111] *See supra* note 26 and accompanying text.

[112] Pan, *supra* note 17. For example, a grocery list can give away as many health insights as a patient record but falls outside the Health Insurance Portability and Accountability Act ("HIPAA"). Angela Chen, *Why It's Time To Rethink the Laws That Keep Our Health Data Private*, VERGE (Jan. 20, 2019, 8:30 AM), https://www.theverge.com/2019/1/29/18197541 [https://perma.cc/9AQE-RDJJ]. See *infra* Part II.A.2 for more information on HIPAA.

[113] Due to the sheer volume of sectoral-based statutes, this Note limits itself to the financial, healthcare, and labor sectors. It does not address other sectors like consumer protection or education, for example.

weaknesses of the regulations discussed in this Part inform the proposal in Part III.

A. Sectoral Regulations

1. *The Financial Sector.* The financial world deals in sensitive personal information: bank balances, account numbers, and credit scores regularly change hands between banks, credit card companies, and other financial institutions.[114] Before 1999, no law required that "financial institutions take any particular measures to fully protect the security and confidentiality of the personal, nonpublic information about their customers."[115] The Gramm-Leach-Bliley Act ("GLBA") created protections for such nonpublic personal information held by financial institutions.[116] Any organization engaging in financial activities, from lending to underwriting,[117] must comply with the GLBA.[118] Specifically, financial institutions must "insure the security and confidentiality of customer records and information," "protect against any anticipated threats or hazards to the security or integrity of such records," and "protect against unauthorized access to or use of such records or information which could result in substantial harm or inconvenience to any customer."[119] The Federal Trade Commission ("FTC") enforces the GLBA.[120]

In practice, the GLBA emphasizes collection and use limitations as well as individual participation. Financial institutions must send an initial privacy notice when first establishing a relationship with a consumer, followed by an annual notice and copy of the privacy policy

[114] *The Gramm-Leach-Bliley Act*, ELEC. PRIV. INFO. CTR. [hereinafter *GLBA*, EPIC], https://epic.org/privacy/glba [https://perma.cc/33H8-WPYU].

[115] H.R. REP. NO. 106-74, at 117–18 (1999).

[116] 15 U.S.C. § 6801.

[117] *See id.* § 6809(3) (defining financial institution as "any institution the business of which is engaging in financial activities"); 12 U.S.C. § 1843(k)(4) (further defining financial activities as "activities that are financial in nature," such as lending, insuring, underwriting, and providing financial advice, among many others).

[118] 15 U.S.C. § 6801.

[119] *Id.*

[120] *Gramm-Leach-Bliley Act*, FTC, https://www.ftc.gov/enforcement/statutes/gramm-leach-bliley-act [https://perma.cc/EQD8-DFCQ] (explaining that 15 U.S.C. § 6801 "requires the FTC, along with the Federal banking agencies and other regulators, to issue regulations ensuring that financial institutions protect the privacy of consumers' personal financial information").

each year the relationship persists.[121] The institution must send additional notices when sharing information with non-affiliated third parties outside of the GLBA's exceptions.[122] Such notices "must explain what non-public personal information is collected, the types of entities with whom the information is shared, how the information is used, and how it is protected."[123] The company must also offer consumers the ability to opt-out of information sharing entirely.[124]

The GLBA has a number of limitations, however. Its emphasis on consumer notice and opt-out mechanisms places the burden on the consumer, not the institution, to protect private data.[125] Given the complexity of the legal language included in the GLBA's notices, this burden is a heavy one.[126] The GLBA also has a narrow focus. It covers only "personally identifiable *financial* information" provided in specific situations[127]—a definition that does not encompass voice data collected via calls to a bank's customer service center. A financial institution would not violate the GLBA by collecting and analyzing voice data to assign loan risk categories.[128]

The Fair Credit Reporting Act ("FCRA") also regulates the financial sector, albeit with a focus on companies "that use data to determine creditworthiness, insurance eligibility, suitability for employment, and to screen tenants."[129] First enacted in 1970, the FCRA imposes limits on data sharing within the credit industry and allows consumers to report errors in consumer reports.[130] The Act follows a three-part model: it "(i) provide[s] notice to consumers of a specific type of data record, (ii) establishe[s] an administrative redress

[121] 15 U.S.C. § 6803; Lisa J. Sotto & Aaron P. Simpson, *United States*, *in* DATA PROTECTION AND PRIVACY IN 26 JURISDICTIONS WORLDWIDE 193 (Rosemary P. Jay ed., 2d ed. 2014), https://www.huntonak.com/images/content/3/3/v3/3351/United-States-GTDT-Data-Protection-and-Privacy-2014.pdf [https://perma.cc/3HY4-LW6Z].

[122] 15 U.S.C. § 6802.

[123] *See* Sotto & Simpson, *supra* note 121, at 193 (explaining 15 U.S.C. §§ 6802 and 6803).

[124] 15 U.S.C. § 6802.

[125] *GLBA*, EPIC, *supra* note 114.

[126] The GLBA runs the danger of creating a rule that provides no real protection: "most privacy and opt-out policies are usually convoluted, confusing, and misleading since they are created by entities whose interests are better served when there is no effective notice." *Id.*

[127] 15 U.S.C. § 6809(4) (emphasis added).

[128] *See* Chen, *supra* note 2 and accompanying text.

[129] *Credit Reporting and Financial Privacy*, FTC (Jan. 2017), https://www.ftc.gov/reports/privacy-data-security-update-2016#credit [https://perma.cc/XV27-DRDB]; 15 U.S.C. § 1681.

[130] 15 U.S.C. § 1681c-1; Boyne, *supra* note 24, at 300.

procedure administered by a government agency, and (iii) define[s] the conditions under which law enforcement could access the data by meeting various standards of proof."[131]

Yet, like the GLBA, the FCRA has a limited reach. It covers only "consumer reports,"[132] or "information by a consumer reporting agency bearing on a consumer's credit worthiness, credit standing, credit capacity, character, general reputation, personal characteristics, or mode of living."[133] The FCRA applies only where a connection to a consumer reporting agency exists. The consumer reporting agency may author the report, a company may use a consumer report created by such an agency, or a company may provide information to power such a report.[134] But if a bank collects information about such a trait without the involvement of a credit reporting agency, the FCRA does not apply. Again, a bank using voice data to assign credit risk would slip through the regulatory cracks.

2. *The Healthcare Sector*. The Health Insurance Portability and Accountability Act ("HIPAA") acts as the foundational health privacy law in the United States.[135] Issued by the Department of Health and Human Services ("HHS"),[136] the HIPAA Privacy Rule establishes privacy standards for protected health information ("PHI").[137] This data category covers information regarding patient health conditions,

[131] Boyne, *supra* note 24, at 300. For more information about the FCRA, see generally *The Fair Credit Reporting Act (FCRA) and the Privacy of Your Credit Report*, ELEC. PRIV. INFO. CTR., https://epic.org/privacy/fcra [https://perma.cc/L882-FSJY], summarizing FCRA's provisions.

[132] *See* DANIEL J. SOLOVE & PAUL M. SCHWARTZ, INFORMATION PRIVACY LAW 757–58 (2021) (explaining that the FCRA's scope turns on its charge to regulate "'any consumer agency' that furnishes a 'consumer report'").

[133] 15 U.S.C. § 1681a(d)(1).

[134] Boyne, *supra* note 24, at 304 fig.1.

[135] Janine Hiller, Matthew S. McMullen, Wade M. Chumney & David L. Baumer, *Privacy and Security in the Implementation of Health Information Technology (Electronic Health Records): U.S. and EU Compared*, 17 B.U. J. SCI. & TECH. L. 1, 11 (2011).

[136] HIPAA's Administrative Simplification Provisions called for Congress to pass legislation protecting individual health data privacy within three years of the bill's passage, or the responsibility to do so would pass to HHS. Health Insurance Portability and Accountability Act, Pub L. No. 104-191, § 264, 110 Stat. 1936, 2033 (1996) (codified at 42 U.S.C. § 1320d-2). Congress failed to accomplish its task, prompting HHS to create the HIPAA Privacy Rule. Stacey A. Tovino, *A Timely Right to Privacy*, 104 IOWA L. REV. 1361, 1368 (2019).

[137] *Summary of the HIPAA Privacy Rule*, HHS, https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations [https://perma.cc/B4U8-UCVD] (last updated July 26, 2013).

health treatment history, and healthcare payments.[138] The HIPAA Privacy Rule kicks in only for PHI held or transmitted by a limited subset of organizations referred to as "covered entities," which includes only health plans, health clearinghouses, and healthcare providers that transmit health information in electronic form.[139]

These covered entities must follow three rules when using or disclosing PHI, each rule more restrictive than the last.[140] First, covered entities "may freely use and disclose PHI without any form of prior permission in order to carry out certain treatment, payment, and health care operations activities, as well as certain public benefit activities."[141] Second, some activities require covered entities inform an individual "in advance of [any] use or disclosure," giving the individual "the opportunity to agree to[,] prohibit[,] or restrict the use or disclosure."[142] Finally, covered entities must "obtain[] . . . a valid authorization" of an individual prior to any use or disclosure of the individual's PHI where the first or second rules do not apply.[143] This third rule acts as the default standard.[144]

The HIPAA Privacy Rule also requires covered entities to proceed cautiously with PHI disclosure and outsourcing. The rule "imposes a general 'minimum necessary' requirement" that limits use and disclosure to only those organizations "required to perform a task."[145] Entities must determine what PHI to allow different types of employees to view and what PHI to release for both routine and non-routine inquiries.[146] Finally, the HIPAA Privacy Rule requires covered entities to create formal contracts with any business associates that "use PHI to perform functions on their behalf."[147] These functions may include "claims processing, data analysis, utilization review, and

[138] 45 C.F.R. § 160.103 (2021).

[139] 45 C.F.R. § 160.102. A healthcare clearinghouse "[p]rocesses or facilitates the processing of health information received from another entity," typically to provide billing, repricing, or information management system services. 45 C.F.R. § 160.103.

[140] Tovino, *supra* note 136, at 1370.

[141] *Id.*; 45 C.F.R. § 164.512(k)(6).

[142] 45 C.F.R. § 164.510.

[143] 45 C.F.R. § 164.508(a)(1).

[144] Tovino, *supra* note 136, at 1371; 45 C.F.R. § 164.508(a)(1).

[145] *Medical Record Privacy*, ELEC. PRIV. INFO. CTR., https://epic.org/privacy/medical/#federalLaw [https://perma.cc/2NDS-4BJS]; 45 C.F.R. §§ 164.502(b), 164.514(d).

[146] *Medical Record Privacy*, *supra* note 145; 45 C.F.R. § 164.504(f)(2)(iii)(A).

[147] *Medical Record Privacy*, *supra* note 145; 45 C.F.R. § 164.504(e)(2).

billing."[148] These requirements, when coupled with the rules outlined above, indicate a reliance on use limitations and individual participation.

Recently, new legislation has tweaked some of HIPAA's provisions; despite these changes, the Privacy Rule still leaves swaths of health data unregulated. The recent Health Information Technology and Economic Clinical Health Act ("HITECH") adds privacy protections to the existing HIPAA framework.[149] HITECH clarifies the application of the HIPAA Privacy Rule to business associates and imposes higher penalties for violations by covered entities and business associates alike.[150] It also requires that covered entities and business associates notify consumers in the case of a data breach.[151] Yet, even HITECH does not stretch HIPAA to cover "health care data generated outside of covered entities and business associates."[152] An electrocardiogram ("EKG") taken by a doctor and recorded in an electronic health record enjoys HIPAA protections, but an EKG taken by an Apple Watch does not.[153] And HIPAA certainly does not protect "the huge volume of data that is not about health at all, but permits inferences about health"[154]—including voice data.

Additional laws govern specific types of health information. At the federal level, the Genetic Information Nondiscrimination Act ("GINA") prevents discrimination based on genetic data in health insurance and employment.[155] At the state level, some states, most notably Illinois, have passed laws regulating biometric data.[156] The Illinois Biometric Information Privacy Act ("BIPA") regulates the collection, retention, disclosure, and destruction of biometric

[148] *Summary of the HIPAA Privacy Rule*, *supra* note 137; 45 C.F.R. § 160.103(4)(i).

[149] 42 U.S.C. §§ 17931–17940.

[150] 42 U.S.C. §§ 17931, 17934; Hiller et al., *supra* note 135, at 13, 18.

[151] 42 U.S.C. § 17932; Hiller et al., *supra* note 135, at 14.

[152] W. Nicholson Price II & I. Glenn Cohen, *Privacy in the Age of Medical Big Data*, 25 NATURE MED. 37, 39 (2019). For example, HIPAA does not cover "health care-related information recorded by life insurance companies." *Id.*

[153] *See id.* ("HIPAA's covered entities, are being supplanted in the health data space by behemoths like Google, Apple, and IBM—all of which operate outside of HIPAA's regime."); Chen, *supra* note 112 ("HIPAA is really about health *care* data more than health data . . . .").

[154] Price & Cohen, *supra* note 152.

[155] Genetic Information Nondiscrimination Act of 2008, Pub. L. No. 110-233, 122 Stat. 881 (codified as amended in scattered sections of 26 U.S.C., 29 U.S.C., and 42 U.S.C.).

[156] For an overview of state legislation on this issue, see generally Sharon Roberg-Perez, *The Future Is Now: Biometric Information and Data Privacy*, 31 ANTITRUST 60 (2017).

identifiers, including voiceprints.[157] Companies must inform individuals of collection and storage policies in writing and receive written consent before they may "collect, capture, purchase, receive through trade, or otherwise obtain" any individual biometric data.[158]

Unfortunately, neither law protects voice insights. GINA covers only genetic information,[159] and BIPA protects the human voice only as an identifier.[160] BIPA only regulates the collection of voice data, not the creation and storage of voice-inferred information.[161] This means the Illinois law fails to address the reality that voice data, while identifying, can also reveal sensitive, personal information about an individual.[162] BIPA's focus on collection limitations may prove an effective safeguard against biometric data misuse, but it does not fix the problem of voice-inferred information. Like HIPAA, GINA and BIPA leave voice insights unaddressed, thus failing to provide suitable privacy safeguards.

3. *Labor and Employment.* Employers increasingly use big data to evaluate prospective and current employees.[163] Historically, employees "enjoy[ed] few privacy rights in the workplace,"[164] but today's workers receive increased privacy protections at both the state

[157] *See* 740 ILL. COMP. STAT. 14/10 (2019) (including voiceprint in the definition of "biometric identifier"). The Illinois General Assembly passed its biometric data law in 2008, noting that "[m]ajor national corporations ha[d] selected the City of Chicago and other locations in this State as pilot testing sites for new applications of biometric-facilitated financial transactions, including finger-scan technologies at grocery stores, gas stations, and school cafeterias." *Id.* 14/5.

[158] *Id.* 14/15.

[159] Even for genetic information, GINA protections fall short—the law does not extend its protections to the life insurance sector. Chen, *supra* note 112. If a long-term insurer finds out a consumer's DNA test predicted early-onset Alzheimer's, "that's information the company can use to change the price of a person's policy or deny them coverage altogether." *Id.*

[160] *See* 740 ILL. COMP. STAT. 14/10 (focusing solely on voiceprints).

[161] *See supra* note 107 and accompanying text.

[162] *See* Andrew McStay, *Emotional AI, Soft Biometrics and the Surveillance of Emotional Life: An Unusual Consensus on Privacy*, BIG DATA & SOC'Y, Jan.–June 2020, at 1, 1 (raising the question of how society should treat "soft biometrics" that can identify bodily traits or emotions without identifying an individual).

[163] *See, e.g.*, Adam S. Forman, Nathaniel M. Glasser & Matthew S. Aibel, *Minimize Risks When Using Big Data Analytics in Hiring*, SHRM (July 12, 2018), https://www.shrm.org/resourcesandtools/legal-and-compliance/employment-law/pages/big-data-analytics-in-hiring.aspx [https://perma.cc/4UUH-NRUV] (noting HR departments increasingly rely on big data analytics). For example, employers might "mine the data of current employees in [a] role [to] find character traits that help define the skills needed to succeed in the role." *Id.*

[164] Boyne, *supra* note 24, at 313.

and federal levels.[165] Companies must comply with many of the data laws discussed above, like HIPAA, GINA, and the FCRA. They must also meet the standards put forth in the Americans with Disabilities Act and the Family and Medical Leave Act.[166] These laws protect employees from discrimination based on personal information, similar to GINA.[167] Some states have further protections. Illinois recently passed the Employee Credit Privacy Act, "which prohibits, with some limited exceptions, inquiries into or obtaining an employee's or applicant's credit history unless there is a specific 'bona fide' reason."[168]

Employers must also comply with the Electronic Communications Privacy Act ("ECPA") when monitoring employee emails and phone calls.[169] But the ECPA allows such monitoring as long as it "is done in the 'ordinary course of business.'"[170] The "ordinary course of business" includes both "monitoring employee e-mail" and "track[ing] the websites visited by their employees."[171] Moreover, the ECPA does not apply where one party consents to surveillance.[172] Often, if the employer "own[s] the email or communications system used by employees, the employees may be deemed to have given [this] consent."[173] And the law does not cover "other forms of monitoring, such as GPS and electronic wearable devices," at all.[174]

Nor does the ECPA prevent employers from using employee voice data to create voice insights. The pre-hire video discussed in the Introduction, for example, receives no ECPA protection. Because applicants must consent to the platform's terms of service to create a video, the ECPA does not apply.[175] Vendors own the videos users

[165] Karin McGinnis, *The Ever Expanding Scope of Employee Privacy Protections*, MOORE & VAN ALLEN (Dec. 2014), https://www.mvalaw.com/news-publications-373.html [https://perma.cc/KGS5-53CU].

[166] *Id.*

[167] *See id.* (noting specifically the Americans with Disabilities Act and the Family and Medical Leave Act as analogs).

[168] *Id.* (quoting 820 ILL. COMP. STAT. 70/10 (2019)).

[169] 18 U.S.C. § 2511; Boyne, *supra* note 24, at 313.

[170] Boyne, *supra* note 24, at 313 (quoting 18 U.S.C. § 2510(5)(a)).

[171] *Id.* at 313–14.

[172] 18 U.S.C. § 2511(2)(c); Richard A. Bales & Katherine V.W. Stone, *The Invisible Web at Work: Artificial Intelligence and Electronic Surveillance in the Workplace*, 41 BERKELEY J. EMP. & LAB. L. 1, 31 (2020).

[173] Bales & Stone, *supra* note 172, at 31.

[174] *Id.*

[175] *See supra* note 172 and accompanying text.

upload; HireVue, one such vendor, acknowledges that it "collects, retains, and stores information" provided by applicants.[176] Individuals cannot simply ask the company to delete their data. No federal data privacy law in the United States guarantees a right to be forgotten.[177] Instead, individuals must choose between forgoing a job opportunity or "creat[ing] a permanent electronic resume . . . that can be neither erased nor challenged."[178]

B. Comprehensive Regulations[179]

1. *Children's Online Privacy Protection Act of 1998.* The United States has had a broad data protection statute since the late 1990s, albeit one limited to children. The Children's Online Privacy Protection Act ("COPPA") "regulates the collection and use of information collected from children under the age of thirteen by Internet websites and mobile apps."[180] It requires companies to gain parental consent prior to obtaining and disclosing children's data.[181] COPPA focuses specifically on safeguarding identifiers, like name (including username), address, telephone number, social security number, persistent identifiers (e.g., IP address or cookie[182]), geolocation, photographs, videos, and audio files.[183] It also includes a catch-all provision that protects any data collected from a child that is later combined with such an identifier.[184]

For all its breadth, COPPA still does not protect children's voice data or voice-inferred information. The statute does not list voice data

[176] Bales & Stone, *supra* note 172, at 33–34.

[177] Brooke Auxier, *Most Americans Support Right To Have Some Personal Info Removed from Online Searches*, PEW RSCH. CTR. (Jan. 27, 2020), https://www.pewresearch.org/fact-tank/2020/01/27/most-americans-support-right-to-have-some-personal-info-removed-from-online-searches [https://perma.cc/CB2V-XCV7] ("[T]he United States has no law or regulatory requirement about removal of personal information from search results or databases.").

[178] Bales & Stone, *supra* note 172, at 33.

[179] *See supra* note 44 and accompanying text.

[180] Boyne, *supra* note 24, at 310; Children's Online Privacy Protection Act of 1998, 15 U.S.C. §§ 6501–6505.

[181] 16 C.F.R. § 312.5 (2021); *see also Children's Online Privacy Protection Act (COPPA)*, ELEC. PRIV. INFO. CTR., https://epic.org/privacy/kids [https://perma.cc/P24Z-RTDD] (reporting the history and operative provisions of COPPA).

[182] *See supra* note 30.

[183] 16 C.F.R. § 312.2.

[184] *Id.*

or voice-inferred information among its enumerated identifiers.[185] Nor does the catch-all provision provide a backstop: because voice insights are created, not collected, they do not implicate COPPA even when combined with an identifier.[186] While FTC commissioners have expressed concerns about the use of predictive analytics on children's data, FTC actions have yet to tackle the issue head on.[187] COPPA appears too narrow to protect children from inferred information like voice insights.

2. *California Consumer Privacy Act.* California recently enacted the California Consumer Privacy Act ("CCPA"), the United States' first non-sectoral data protection statute.[188] The CCPA creates legal protections that "follow personal data, regardless of whether an individual has a direct relationship with the regulated company."[189] It has four major provisions: (1) the right to know what personal data a company has collected and disclosed,[190] (2) the right to opt-out of having companies sell personal data to third parties,[191] (3) the right to

[185] *Id.*

[186] *See* Hideyuki Matsumi, *Predictions and Privacy: Should There Be Rules About Using Personal Data To Forecast the Future?*, 48 CUMB. L. REV. 149, 178–79 (2018) (arguing that predictive information likely does not qualify as information collected from children, even when combined with identifiers).

[187] *Id.* at 179 n.204.

[188] California Consumer Privacy Act of 2018, CAL. CIV. CODE §§ 1798.100–1798.199 (West 2020); Bales & Stone, *supra* note 172, at 32 ("The CCPA is the first omnibus privacy regulation in the United States . . . ."). Although a state law, the CCPA serves as a harbinger for the rest of the country's data protection schemas. For one thing, "most major companies do business in the state and, as a result, are impacted" by the law's privacy mandates. Jeff John Roberts, *New California Law Giving Consumers Control Over Their Data Sets Off a Scramble*, FORTUNE (Dec. 18, 2019, 6:30 AM), https://fortune.com/2019/12/18/california-consumer-privacy-act-data-nationwide [https://perma.cc/L8S4-NNC4]. For another, the proliferation of state-specific regulation has increased the calls for a single federal law "as the business community howls at the prospect of complying with a patchwork of state requirements." Gilad Edelman, *California's Privacy Law Goes Into Effect Today. Now What?*, WIRED (Jan. 1, 2020, 7:00 AM), https://www.wired.com/story/ccpa-guide-california-privacy-law-takes-effect [https://perma.cc/T4MR-QDW7].

[189] Anupam Chander, Margot E. Kaminski & William McGeveran, *Catalyzing Privacy Law*, 105 MINN. L. REV. 1733, 1749 (2021).

[190] *Id.* at 1751–52; CIV. § 1798.100.

[191] Chander et al., *supra* note 189, at 1753; CIV. § 1798.120.

have a company delete personal data,[192] and (4) the right to equal treatment regardless of the invocation of rights under the CCPA.[193]

Although the CCPA shares some similarities with European data law, it does not copy it exactly.[194] Like the GDPR, the CCPA emphasizes "transparency and accountability from companies and control for data subjects."[195] But the CCPA does not stray far from the U.S. "notice and consent" model. It still puts much of the onus on consumers to control how their personal data is used.[196] It also lacks the broad coverage of the GDPR,[197] applying only to businesses that fit its complex requirements.[198] Still, the GDPR can provide helpful insight about how the CCPA may fare with voice-inferred information, given the European statute's longer tenure.

The GDPR and the CCPA share a gap: both laws focus on information a company has *collected* about a consumer, not *created* about a consumer.[199] Under the EU law, data controllers[200] must notify users about "the categories of personal data collected, intended purposes of processing, recipients or categories of third-party recipients, the data controller's or third party's legitimate interests

[192] Chander et al., *supra* note 189, at 1754–55; CIV. § 1798.105. But note that this right does not extend to third parties that do not collect the data directly from the consumer. Chander et al., *supra* note 189, at 1754.

[193] Chander et al., *supra* note 189, at 1753; CIV. § 1798.125.

[194] Hartzog & Richards, *supra* note 42, at 1711.

[195] *See id.* at 1693 (discussing similarities between states' proposed data protection legislation and the GDPR as putting further pressure on Congress to pass a federal data law).

[196] *See id.* at 1711–12 (noting that several rights under the CCPA must be exercised by consumers and are not self-effectuating).

[197] *See* Chander et al., *supra* note 189, at 1758 (comparing the CCPA's scope to the GDPR, which "covers anyone that processes personal data, including not only companies but also individuals, nonprofit organizations, and governments").

[198] *See id.* at 1758 & n.161 (describing the CCPA's "overlapping requirements related to [the company's] size or the extent of their involvement in personal data trade," including revenue and customer base requirements). For more information, see CIV. § 1798.140(c).

[199] *Cf.* Mary T. Costigan, *CPRA Series: Sensitive Personal Information*, JACKSONLEWIS (Dec. 14, 2020), https://www.workplaceprivacyreport.com/2020/12/articles/california-consumer-privacy-act/cpra-series-sensitive-personal-information [https://perma.cc/MMZ4-VKXX] (explaining how these laws police the collection of consumer information).

[200] The GDPR defines a "controller" as a "natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data." GDPR, *supra* note 25, at 33. Processors then process the data for the controller, who can subcontract some of the work to a subprocessor. *Id.* at 33, 49. Confused? Think of buying your favorite brand of all-purpose flour to feed a sourdough starter: the controller is King Arthur Flour, which operates the mill; the processor grinds the wheat into flour; and the subprocessor harvests the wheat to take to the mill.

justifying processing . . . , and 'from which source the personal data originate.'"[201] Yet, this requirement only applies where a controller obtains data from a data subject or third party.[202] Where the data controller creates inferred data itself, "notification duties will never be triggered."[203] European Economic Area citizens and residents have the right to access their data, but this right provides little help: individuals would still somehow need to know the data existed and which controller held it before making any data request.[204] The GDPR fails to address inferred information like voice insights. The CCPA will likely fare no better.

Broader laws like COPPA and the CCPA give voice insights no more protection than sector-specific data privacy laws. Use and collection limitations can only go so far in the age of big data analytics: voice insights need a different solution.

III. TOWARDS A PURPOSE-BASED PRIVACY APPROACH FOR VOICE DATA

Faced with a veritable alphabet soup of privacy laws, support for a federal comprehensive privacy law has grown.[205] Any such legislation should look forward, not backward: it should confront the rise of big data analytics and voice insights head on. To do so, the United States must look beyond collection and use limitation principles to embrace purpose specification.

Applying collection and use limitation principles to voice data makes for an awkward fit.[206] Policymakers cannot just ban voice data collection outright; such a blanket prohibition would clash with other federal regulations, like those that require financial firms to monitor and record customer calls.[207] Neither can policymakers rely on use-based consent. If you agree to let a company use your voice data in one

[201] Wachter & Mittelstadt, *supra* note 16, at 544.

[202] *Id.* at 545.

[203] *Id.*

[204] *Id.* at 545–46.

[205] *See supra* notes 43–44 and accompanying text.

[206] See *supra* notes 39–42 for definitions of the use and collection limitation principles.

[207] *See 3170. Tape Recording of Registered Persons by Certain Firms*, FINRA, https://www.finra.org/rules-guidance/rulebooks/finra-rules/3170 [https://perma.cc/L74B-9QXK] (requiring members to "establish, maintain, and enforce special written procedures for supervising the telemarketing activities of all of its registered persons"). Despite its unworkability for the general public, this approach may make more sense in the context of children's voice data.

context, the company can then use that data for any purpose it can imagine.[208] Both collection and use limitations still play an important role in ensuring data privacy. But these principles need support from a more powerful mechanism: purpose specification.

The purpose specification principle affords voice insights better protections.[209] Under this principle, a company that initially collects voice data for quality assurance purposes can only use the collected data to that end.[210] The company cannot simply decide to later use the recorded data for loan risk purposes (or hiring purposes, or mental health diagnosis purposes). To use the data for a new purpose, the company would need to gain new customer consent. This protection extends to data companies create through predictive analytics: a company would need consumer agreement to use previously collected data to create new insights.[211] Companies could no longer simply plug a voice data point into an algorithm to see what information they can infer.

Some purpose changes will still need to take place, of course. Any data protection law will need a mechanism that allows reasonable changes to purpose while leaving privacy protections intact. After all, legitimate use cases may arise after a company initially collects user data. Innovation, too, increasingly relies on machine learning and analytics.[212] Parts III.A and III.B discuss two purpose-based protections that balance data privacy with innovation. While each can work as a standalone solution, they can also function together as complementary protections.[213]

A. *Require Meaningful, Purpose-Based Consent*

Unless the United States completely overhauls its data regulatory system, the principles of notice and consent will likely remain an integral part of any future data privacy law. Working within this framework, then, can provide a pragmatic and achievable path towards

[208] *See supra* notes 39–40 and accompanying text (describing the use limitation principle's sole focus on consent for collection, rather than application of the data).

[209] *See supra* notes 37–38 and accompanying text.

[210] *See supra* notes 37–38 and accompanying text.

[211] *See supra* notes 37–38 and accompanying text.

[212] See *supra* Part I for a discussion of the current data technology, analytics, and innovation.

[213] This Note suggests these solutions have particular applicability in the voice insights context. They may also prove helpful in addressing other privacy concerns, but this exceeds the scope of the Note.

protecting voice insights. The United States should draw on its notice and consent roots to require meaningful consumer consent for any data purpose change.[214]

Meaningful consent would require companies to specify the purpose for which they collect any data, including voice data; should this purpose change at a later date, the company would then need to obtain new consent from the consumer. To qualify as meaningful, consumer consent would need to be affirmative and explicit. An email or website banner simply declaring the company's terms have changed, without more, would not pass this standard for two reasons. First, the notice lacks any details about the change. To qualify as explicit, any notice prompting a user's consent must include information about the change and its effect on user data. Second, the notice lacks a call to action; it relies only on implied consent "inferred from the action or inaction of the individual."[215] Meaningful consent would require an affirmative opt-in from consumers. Companies could not simply "rely on silence, inactivity, default settings, pre-ticked boxes or . . . general terms and conditions, or seek to take advantage of inertia, inattention or default bias in any other way."[216]

By requiring affirmative, explicit consent, data privacy becomes the default setting. Putting individuals in charge of their own data can empower, but it can also overwhelm. People already face an onslaught of consent requests: "[m]obile apps can ask users for over 200 permissions and even the average app asks for about five."[217] This constant barrage desensitizes consumers to data requests, leading people to ignore even the most obtrusive notifications.[218] An opt-out consent model allows companies to capitalize on consumers' limited

[214] See *supra* Part II for a discussion of the current U.S. approach to data protection.

[215] *Consent*, IAPP, https://iapp.org/resources/article/consent-2 [https://perma.cc/SP9E-CMED].

[216] *What Is Valid Consent?*, INFO. COMM'R'S OFF., https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/consent/what-is-valid-consent [https://perma.cc/VX8D-LVZY].

[217] Woodrow Hartzog, *The Case Against Idealising Control*, 4 EUR. DATA PROT. L. REV. 423, 429 (2018) [hereinafter Hartzog, *Idealising Control*].

[218] *Id.* For example, the GDPR drove an increase of cookie notices across the internet, aimed at giving consumers control over their web tracking data. Matt Burgess, *We Need To Fix GDPR's Biggest Failure: Broken Cookie Notices*, WIRED (May 28, 2020, 6:00 AM), https://www.wired.co.uk/article/gdpr-cookie-consent-eprivacy [https://perma.cc/38AC-RYPL]. A year or two into the change, the tool has proved largely ineffective: people simply click through the notifications to get rid of the distraction on the screen. *Id.*

capacity to sort through every data request they receive. Worse, design can nudge people to accept privacy requests through "dark patterns" that exploit human psychology.[219]

Requiring meaningful consent combats this decision fatigue. A consumer need not worry about ignoring a terms of service change. Without the consumer's affirmative consent, the company's metaphorical hands remain tied: it cannot simply assume that silence signals acceptance of the change. Thus, meaningful consent minimizes the threat of inadvertent acceptance of terms of service against the consumer's interest. Consumers remain empowered to share their data should they so desire, but they need not act on every request or notification to protect their data. Nor do they need to self-censor their spoken conversations. No company could generate voice insights without explicit, affirmative consent.[220] Meaningful consent protects data and voice data alike, allowing people to speak freely without weighing the privacy implications of each word.

B. Implement a Data Review Board

Leaving the realm of notice and consent, policymakers should consider creating a data governance body similar to the Institutional Review Boards ("IRBs") that monitor clinical research. Created in 1974, IRBs today "function as a kind of ethics committee," making sure "the rights and welfare of research subjects" remain protected.[221] An IRB sits within every federally funded university or organization

[219] Burgess, *supra* note 218. Platforms capitalize on humans' "built-in tendenc[y] to prefer shiny, colourful buttons and ignore dull, grey ones" to gain preferred results. Hartzog, *Idealising Control*, *supra* note 217, at 427.

[220] Companies may voice concern over the cost of compliance with such a rule: California has estimated that initial compliance with the CCPA has cost $55 billion. OFF. OF THE ATT'Y GEN., STATE OF CAL. DEP'T OF JUST., STANDARDIZED REGULATORY IMPACT ASSESSMENT: CALIFORNIA CONSUMER PRIVACY ACT OF 2018 REGULATIONS 11 (2019). These costs, however, are at least partially offset by the value of the personal data protected: over $20 billion annually in California alone. *Id.* at 13, 15. And, given the increase in cybercrime—data breaches exposed 4.1 billion private records in just the first six months of 2019—protecting personal data should not be undervalued. Davey Winder, *Data Breaches Expose 4.1 Billion Records in First Six Months of 2019*, FORBES (Aug. 20, 2019, 6:31 AM), https://www.forbes.com/sites/daveywinder/2019/08/20/data-breaches-expose-41-billion-records-in-first-six-months-of-2019/?sh=4b7b25a9bd54 [https://perma.cc/ZJ33-PB37].

[221] ROBERT J. AMDUR & ELIZABETH A. BANKERT, INSTITUTIONAL REVIEW BOARD MEMBER HANDBOOK 5, 16 (3d ed. 2011). Prior to the establishment of the IRB, a series of atrocities plagued U.S. human subject research. *See id.* at 7–16 (cataloging unethical biomedical and social science research, including the notorious Tuskegee Syphilis Study).

conducting such research,[222] and it must approve any research involving human subjects before the project begins.[223] Researchers must provide "a full description of the proposed project," along with information about the project's materials, recruitment strategy, and consent form.[224] The researchers must also describe "how the subjects' confidentiality will be maintained."[225] Using this information, the IRB then determines whether the project adequately protects its participants.[226] If the IRB perceives risks, it can request specific changes or revisions, or even reject the project entirely.[227]

IRBs provide a helpful analog for data privacy oversight because of the similarities between human subject research and predictive analytics. Like predictive analytics, human research often involves "us[ing], study[ing], analyz[ing], or generat[ing] identifiable information."[228] And, like inferred information, these identifiers include data that can reveal identity "through deductive disclosure (e.g., a combination of unique characteristics, such as a student's gender, year in school, major, and athletic affiliation)."[229] Yet, unlike federally funded universities and organizations, companies face no similar oversight.[230] Instead, the onus is on individuals to protect their data from company overreach.

Creating a data IRB would shift the burden off the consumer and onto a board of experts trained to recognize privacy threats. This expertise would prove particularly valuable in the context of voice data privacy. Given the unintuitive nature of voice insights, an audio clip can pose more privacy threats than a layperson might recognize.

[222] *Frequently Asked Questions About Institutional Review Boards*, AM. PSYCH. ASS'N (Sept. 2017), https://www.apa.org/advocacy/research/defending-research/review-boards [https://perma.cc/94Q8-JCCR].

[223] *Id.*

[224] *Id.*

[225] *Id.*

[226] *Id.*

[227] *Id.*

[228] *Campus Institutional Research Board: Before You Begin*, DUKE UNIV. CAMPUS INSTITUTIONAL REV. BD., https://campusirb.duke.edu [https://perma.cc/L4NZ-FKBQ]. Indeed, IRBs frequently review projects involving "secondary analysis of a data set gathered for another purpose," similar to the work done when generating voice insights. *IRB Frequently Asked Questions*, UCI OFF. OF RSCH., https://www.research.uci.edu/compliance/human-research-protections/researchers/irb-faqs.html#Does [https://perma.cc/7KU3-XGRK].

[229] DUKE UNIV. CAMPUS INSTITUTIONAL REV. BD., *supra* note 228.

[230] *See supra* note 222 and accompanying text.

Individuals might consent to voice-inferred information without realizing the true privacy ramifications of the decision. Data IRBs would help correct this knowledge imbalance by allowing only beneficent voice projects to move forward. Consumers would finally receive the same protections research participants already enjoy.

Companies may protest that a data review process will inhibit their ability to "operate at speed and scale, protect trade secrets, and satisfy investors."[231] After all, even traditional IRBs have faced criticism for being "plodding or skewed."[232] But many companies already incorporate similar review processes into their own internal decision making. To receive corporate funding, projects must create "a business case and a plan with a fixed scope, schedule and cost" for upper-level management approval.[233] Companies invest in this time-consuming process because of its positive impact on financial health.[234] Data reviews can bring similar long-term benefits.[235] For example, "[data IRBs] could help unearth and head off media fiascos before they materialize," "increase regulatory certainty," and "add a measure of legitimacy to the study of consumers for profit."[236] Indeed, companies like Facebook and Palantir have already created their own data review boards to minimize privacy impacts and ensure algorithmic fairness.[237] As predictive analytics continues to grow, data oversight will become an asset, not a liability.

Data IRBs will likely differ from traditional IRBs in meaningful ways. Rather than follow the same standards as clinical IRBs, data review boards should follow privacy-specific guidelines like the

[231] Ryan Calo, *Consumer Subject Review Boards: A Thought Experiment*, 66 STAN. L. REV. ONLINE 97, 101 (2013), https://review.law.stanford.edu/wp-content/uploads/sites/3/2016/08/Calo.pdf [https://perma.cc/M5BZ-V7NB].

[232] *Id.*

[233] Hakan Altintepe, *Product Funding and the Burden of Agility*, CIO (June 21, 2019, 5:55 AM), https://www.cio.com/article/3404456 [https://perma.cc/7CZY-J7EB].

[234] *See* Brian Herman & Jay M. Siegelaub, *Is This Really Worth the Effort? The Need for a Business Case*, PMI (Oct. 13, 2009), https://www.pmi.org/learning/library/need-business-case-6730 [https://perma.cc/SYF8-43L6] (explaining that business cases provide "a more rational and effective means of allocating [] limited resources").

[235] *See* Calo, *supra* note 231, at 102 (identifying how companies can gain public legitimacy and regulatory certainty from such reviews, among other benefits).

[236] *Id.*

[237] Sam Shead, *Facebook Reportedly Has a Dedicated AI Ethics Team*, FORBES (May 3, 2018, 5:38 PM), https://www.forbes.com/sites/samshead/2018/05/03/facebook-reportedly-has-a-dedicated-ai-ethics-team [https://perma.cc/7CAB-ETUK]; *Privacy & Civil Liberties Engineering: Advisors*, PALANTIR, https://www.palantir.com/pcl [https://perma.cc/LKR6-6LKG].

FIPPs.[238] And, unlike the current IRBs that sit within organizations, data IRBs should remain independent from the companies they monitor to ensure objectivity. Determining the precise mechanism for a data IRB agency exceeds the scope of this Note.[239] But, two options deserve further exploration.

First, the FTC could absorb a data IRB function under its unfair and deceptive trade practices umbrella.[240] The FTC has become the de facto privacy regulator in the United States, making it an obvious home for such a function.[241] The agency has successfully absorbed new privacy functions in the past.[242] To do so again, however, policymakers will likely need to provide the FTC more resources dedicated specifically to privacy.[243] The FTC has a "broad mission in competition and consumer protection"; of its one thousand total staff, "no more than 50 are tasked with privacy."[244] As a result, the agency only "announces about 15 [to] 20 Section 5 enforcement settlements per year."[245] Adding a data IRB component would thus require investment in the FTC.

Second, Congress could create a technology-specific agency to handle the data review process. Such an idea is not unprecedented. California will soon establish its own Privacy Protection Agency to take over CCPA enforcement,[246] and countries like Canada and New

[238] *See supra* notes 31–38 and accompanying text.

[239] For more information about data IRBs, see generally Jules Polonetsky, Omer Tene & Joseph Jerome, *Beyond the Common Rule: Ethical Structures for Data Research in Non-Academic Settings*, 13 COLO. TECH. L.J. 333 (2015), or Calo, *supra* note 231.

[240] *A Brief Overview of the Federal Trade Commission's Investigative, Law Enforcement, and Rulemaking Authority*, FTC (May 2021), https://www.ftc.gov/about-ftc/what-we-do/enforcement-authority [https://perma.cc/YUJ7-JGC8].

[241] *See* Daniel J. Solove & Woodrow Hartzog, *The FTC and the New Common Law of Privacy*, 114 COLUM. L. REV. 583, 598–606 (2014) (describing the FTC's rise as regulator and enforcer in the privacy space).

[242] For example, the FTC became the enforcement authority for the Safe Harbor Agreement between the United States and Europe in 2000. *Id.* at 603–04; *see also* Commission Decision 2000/520/EC, 2000 O.J. (L 215) 8–9, 33–38 (describing the FTC's enforcement role).

[243] Chris Jay Hoofnagle, Woodrow Hartzog & Daniel J. Solove, *The FTC Can Rise to the Privacy Challenge, but Not Without Help from Congress*, BROOKINGS (Aug. 8, 2019), https://www.brookings.edu/blog/techtank/2019/08/08/the-ftc-can-rise-to-the-privacy-challenge-but-not-without-help-from-congress [https://perma.cc/V9RS-AHFS].

[244] *Id.*

[245] *Id.*

[246] Edward S. Chang, Jennifer C. Everett, Daniel J. McLoon, Mauricio F. Paez, Jeff Rabkin, Lisa M. Ropple & John A. Vogt, *California Voters Adopt the California Privacy Rights Act*, JONES

Zealand employ national privacy commissioners.[247] While individual state agencies could each develop their own data IRBs, a federal privacy agency could also coordinate such review activities. This solution would impose a cost on taxpayers.[248] But a federal privacy statute, enforced by a federal privacy agency, would also alleviate the cost and headache of complying with the different state privacy laws that keep cropping up.[249]

Of course, any data IRB need not function alone. A data oversight board will likely work best when paired with a meaningful consent requirement. The data IRB would act as a gatekeeper, ensuring that any new data analytics project has rigorous privacy protections in place. But consumers would still retain autonomy. Rather than outsource control of personal data entirely to a bureaucratic body, individuals could still choose when to opt-in to data requests from trusted organizations. Together, these solutions would guarantee all data, including voice-inferred information, has sufficient protection.

CONCLUSION

The United States has reached a tipping point in data privacy, driven by increasingly powerful technology and a fragmented regulatory landscape. Imposing purpose specifications in a new federal data privacy law will protect consumers from privacy threats posed by the growing use of voice insights. But privacy protections need not stymy technological advancement. Policymakers and companies can ensure the health of U.S. innovation by allowing purpose changes with meaningful consent and data IRB approval. With such protections in

DAY (Nov. 2020), https://www.jonesday.com/en/insights/2020/11/california-voters-approve-cpra [https://perma.cc/6VE4-PS4D].

[247] OFF. OF THE PRIV. COMM'R OF CAN., https://www.priv.gc.ca/en [https://perma.cc/MB5S-UCCC]; PRIV. COMM'R, https://www.privacy.org.nz [https://perma.cc/8Q92-MBFA].

[248] Alan McQuinn & Daniel Castro, *The Costs of an Unnecessarily Stringent Federal Data Privacy Law*, REAL CLEAR PUB. AFFS., https://www.realclearpublicaffairs.com/public_affairs/2019/08/12/the_costs_of_an_unnecessarily_stringent_federal_data_privacy_law_18753.html [https://perma.cc/396V-8QQC] ("Federal legislation similar to the privacy laws in Europe or California could cost the U.S. economy approximately $122 billion per year, or $483 per U.S. adult.").

[249] *See* Michael Beckerman, Opinion, *Americans Will Pay a Price for State Privacy Laws*, N.Y. TIMES (Oct. 14, 2019), https://www.nytimes.com/2019/10/14/opinion/state-privacy-laws.html [https://perma.cc/5T9E-CZDH] (arguing that "[f]ailure to pass national standards will harm the American economy" because of the cost and difficulty of complying with state-level data privacy laws).

place, consumers would no longer need to fear the power of their own voice.