Synthetic Data's Transformative Role in Foundational Speech Models
31 August 2024, KOS, Greece

Navigating the United States Legislative Landscape on Voice Privacy: Existing Laws, Proposed Bills, Protection for Children, and Synthetic Data for AI

Satwik Dutta, John H.L. Hansen

Center for Robust Speech Systems (CRSS), The University of Texas at Dallas, USA
satwik.dutta@utdallas.edu, john.hansen@utdallas.edu

Abstract

Privacy is a hot topic for policymakers across the globe, including the United States. Evolving advances in AI and emerging concerns about the misuse of personal data have pushed policymakers to draft legislation on trustworthy AI and privacy protection for its citizens. This paper presents the state of the privacy legislation at the U.S. Congress and outlines how voice data is considered as part of the legislation definition. This paper also reviews additional privacy protection for children. This paper presents a holistic review of enacted and proposed privacy laws, and consideration for voice data, including guidelines for processing children's data, in those laws across the fifty U.S. states. As a groundbreaking alternative to actual human data, ethically generated synthetic data allows much flexibility to keep AI innovation in progress. Given the consideration of synthetic data in AI legislation by policymakers to be relatively new, as compared to that of privacy laws, this paper reviews regulatory considerations for synthetic data.

Index Terms: voice privacy, privacy laws, children's privacy, synthetic data, AI legislation.

1. Introduction

Human voice or speech contains very personal information about a speaker and therefore, it is important to safeguard voice or audio recordings of a speaker from misuse. Guidelines on the collection, storage, and use of any individual's personal data, as collected by any business (such as a company, operator, or service provider), need to comply with the privacy policies as set forward by the local, state, and federal agencies and government. Given numerous recent instances of violation of consumer privacy as well as rapidly evolving Artificial Intelligence (AI) technology which is now available to scammers, lawmakers across U.S. and the world have placed the topic of data privacy on the center stage. Countries and diplomatic organizations across the globe are drafting and implementing AI governance legislation and policies. The thrust for legislation on AI in the United States (U.S.) has been both from the executive branch i.e. the office of the U.S. President or the White House, and the legislative branch i.e. U.S. Congress - collectively by the Senate and the House. Privacy has been a key focus for lawmakers in drafting AI legislation. However, many policymakers agree that the legislation on AI and data privacy would be interlinked.

In 2022, the White House Office of Science and Technology Policy released the *Blueprint for the AI Bill of Rights* [1] with outlining principles and good practices to design, use, and deploy AI systems for protecting civil rights, civil liberties, and privacy of the citizens. This was followed by the *Executive Order on Safe, Secure, and Trustworthy AI* [2] by the President in 2023. This extensive executive order covered safety and security of AI technology, promotion of innovation and competition, support for workforce, protection of rights and privacy, including actions for several federal agencies such as the National Institute of Standards and Technology or NIST. On May 15, 2024, a *Roadmap for Artificial Intelligence Policy in the United States Senate* [3] was released by the Bipartisan Senate Artificial Intelligence (AI) Working Group. This roadmap highlights various policy priorities including funding for AI innovation, enforcement of existing laws for AI, impact of AI on workforce, enhancing national security, addressing challenges posed by deepfakes, and support for higher education research and development on AI. This roadmap also prioritizes policies on establishing a strong comprehensive federal data privacy framework.

Motivated by many recent U.S. legislation on privacy protection, this paper aims to give an overview and current state of both the federal and state privacy policies across the U.S. A holistic review of these privacy legislation also highlights how voice fits in the legislative definition. Children's data being sensitive [4, 5, 6], a review of protection for children's privacy legislation is also presented in this paper. As compared to privacy laws, AI legislation is new, and many U.S. states are considering additional AI legislation on top of the privacy laws. Some of these AI legislation also consider defining synthetic data and voice generation. An illustrative timeline of major legislative actions for AI and privacy is shown in Fig.1. This paper is structured as follows: a newly proposed national privacy legislation for the U.S. is discussed in Sec.2, followed by privacy protection guidelines for children in Sec.3. This is followed by a holistic review of the privacy laws across all the U.S. states in Sec.4 and laws on synthetic data for AI in Sec.5. Finally, we conclude this paper in Sec.6.

2. American Privacy Rights Act of 2024

The *American Privacy Rights Act of 2024* (APRA) [7, 8] was proposed in April 2024 by two prominent members of the U.S. Congress who serve as Chairs of the House Committee on Energy and Commerce and Senate Committee on Commerce, Science and Transportation. As the lawmakers mentioned, this proposed legislation is "bipartisan" (support from lawmakers from both political parties) and "bicameral" (support from lawmakers from both chambers of the U.S. Congress). The primary goal of APRA is to establish a nationwide comprehensive data privacy and security standard for all U.S. citizens including children. APRA includes several actions such as (1) guidelines on prohibitions on consumer personal data use, including sensitive and biometric data of consumers, (2) consumer rights, including access, correction, and deletion, on covered data (covered

10.21437/SynData4GenAI.2024-19

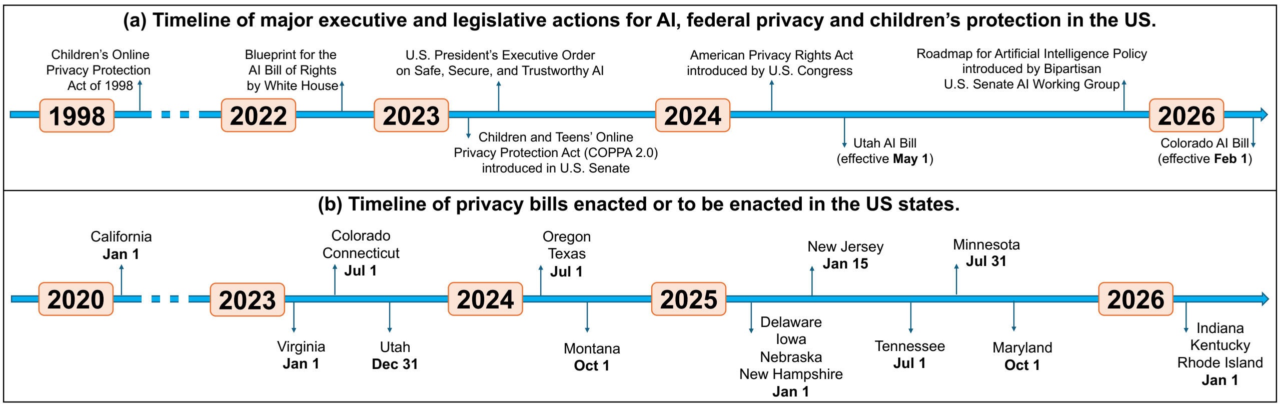
(a) Timeline of major executive and legislative actions for AI, federal privacy and children's protection in the US.
Children's Online Privacy Protection Act of 1998
1998
2022
Blueprint for the AI Bill of Rights by White House
2023
U.S. President's Executive Order on Safe, Secure, and Trustworthy AI
Children and Teens' Online Privacy Protection Act (COPPA 2.0) introduced in U.S. Senate
2024
American Privacy Rights Act introduced by U.S. Congress
Utah AI Bill (effective May 1)
Roadmap for Artificial Intelligence Policy introduced by Bipartisan U.S. Senate AI Working Group
2026
Colorado AI Bill (effective Feb 1)
(b) Timeline of privacy bills enacted or to be enacted in the US states.
2020
California Jan 1
2023
Virginia Jan 1
Colorado Connecticut Jul 1
Utah Dec 31
2024
Oregon Texas Jul 1
Montana Oct 1
2025
Delaware Iowa Nebraska New Hampshire Jan 1
New Jersey Jan 15
Tennessee Jul 1
Minnesota Jul 31
Maryland Oct 1
2026
Indiana Kentucky Rhode Island Jan 1

Figure 1: *(a) Timeline of major executive and legislative actions for AI, federal privacy and children's protection in the US (as of Jun 24, 2024). (b) Timeline of privacy bills enacted or to be enacted in the US states (as of Jun 24, 2024).*

data is any data which identifies or is linked or reasonably linkable to an individual), (3) policies for design by privacy and transparency, (4) opt-out privilege of consumers, and (5) enforcement of internal data security and privacy guardrails by businesses.

This legislation primarily includes all businesses subject to the U.S. Federal Trade Commission's (FTC) authority, common carriers, and non-profits, except small businesses (<USD 40 million in annual revenue, <200,000 consumers). This also includes additional obligations for high-impact social media companies, data holders, and data brokers. After its original draft was released in April, APRA has been discussed and edited twice with an expected markup (lawmakers offer and vote on amendments) scheduled for June 27, 2024, by the House Committee on Energy and Commerce. However, there are still multiple concerns with this legislation.

APRA considers "**voice prints**" as biometric information, which is defined as any data that could directly or indirectly help to identify an individual and that could be generated through unique characteristics of an individual like biological, physical, or physiological. Biometric information is considered as "sensitive covered data". There are guidelines regarding the collection, processing, retention, or transfer of biometric information, and the need for consent from consumers. An "**audio recording**" or any data derived from such a recording is **excluded** from the definition of biometric information unless it could be used to identify an individual. Private communications of an individual such as "voicemails", "voice or video communications", or any related information regarding its transmission and private "audio recordings" are also considered as "sensitive covered data". Additional protections are outlined regarding the transfer of sensitive covered data.

3. Children's Privacy in the US

3.1. 1998 Children's Online Privacy Protection Act

The *Children's Online Privacy Protection Act of 1998* [9], COPPA, was enacted to prohibit unfair collection and use of personal information of children under the age of 13 on the web. In 2013, COPPA was amended to extend the definition of "website or online service directed to children", to expand the definition of "personal information" to include **audio file** where such file contains a child's "voice" (including photo, video), and acceptable methods for verifiable parental consent.

3.2. 2024 Children & Teens' Online Privacy Protection Act

The Children and Teens' Online Privacy Protection Act of 2024 [10], COPPA 2.0, was recently proposed in the U.S. Congress as an update to COPPA. This particular legislation has gained significant momentum since Feb 2024, both at the U.S. Senate and House. It has also been discussed in the House Committee on Energy and Commerce along with other bills such as KOSA (Kids Online Safety Act). COPPA 2.0 would also extend protection to teens between 12 to 17 years in age. Several updates have been proposed in the new COPPA 2.0 legislation, including extension of the definition of personal information, website/online service providers, consent, data retention, advertising, and use for educational technology.

Particularly considering voice, the definition of personal information has been modified for one item and a new item added: (1) "*an **audio** file where such file contains a specific child's or teen's **voice***", and (2) "*information generated from the measurement or technological processing of an individual's biological, physical, or physiological characteristics that is used to identify an individual, including **voice prints***". COPPA 2.0 also adds a specific **exclusion section for audio files**. Audio files are excluded from being considered as personal information if the service or operator: (1) does not request for personal identifiable information, (2) clearly states the collection, use, and deletion policy in its privacy policy, (3) only uses the audio file for the intent or task for which it was collected, (4) maintains the audio file to perform the intent or task, and deletes the audio file immediately without any other use before deletion. These guidelines suggest that **the audio files could only be used for providing a service, and not for any innovation of the product or service**. There is currently no legislative text on de-identification or use of de-identified data. As of May 2024, COPPA 2.0 has been added as a section under APRA. The primary goal for COPPA 2.0 is to bring online data privacy protection for children and teens to the 21st century.

4. State-level Privacy Regulations in the US

Out of all the 50 U.S. states, California was the first state to enact a strict consumer privacy law in 2020 - the *California Consumer Privacy Act of 2018* (CCPA) [11]. Additional privacy protections were added to CCPA in 2020 and amended by the *California Privacy Rights Act of 2020*. Founded in 2020 by CCPA, the California Privacy Protection Agency began updating existing laws and adopting new legislation in 2022. The action on state privacy bills was followed by other states such as: Virginia, Colorado, Connecticut, Utah, and many more. 19 U.S. states have already enacted new privacy legislation. Legislation on privacy has been proposed and is in progress in 13 states, while the action has already failed in 2 states. To date, no privacy bill has been proposed in the remaining 16 U.S. states. An illustrative map of the current status of privacy laws across all U.S. states is shown in Fig.2(a).

Particularly for legislative text on "**voice**", based on our review, most states consider "**voice prints**" as biometric information similar to ARPA. While only California [11] and Illinois [12] add "**voice recording**" in the definition of biometric information, most states exclude "**audio recording**" or any data derived from such a recording as biometric information unless it could be used to identify an individual. Delaware [13] and Pennsylvania [14] also add another specific exclusion[1] in their "biometric information" definition which could refer to raw audio when converted to an array or acoustic features. Most states also define a broader "sensitive data" which includes personal and biometric data (or information). The legislative text for Ohio [15] is very different from all other states, not explicitly defining biometric or sensitive data or including voice or audio recordings. Infographics showing which states include voice in the definition of biometric information and which states add a broader definition of sensitive data (including biometric and personal) are shown in Fig.2(b,c).

Many states also add additional protection for children's privacy by prohibiting the processing of sensitive (or personal) data of children with verifiable consent as shown in Fig.2(d). While most states comply with other state and federal regulations, some explicitly mention and comply with COPPA, as shown in Fig.2(e). It should be noted that California [11] and Illinois [12] have even stricter child privacy regulations as compared to the enacted version of COPPA. While most states consider the protection of child privacy rights below 13 years, few keep the limit up to 16 or 17 years, as shown in Fig.2(f).

5. Regulations on Synthetic Data for AI

Utah [16] is the first U.S. state to enact a comprehensive AI governance law that went into effect on May 1, 2024. Similar to its privacy law [17], "**voice**" is defined in the AI legislative text. In this legislation, "synthetic data" is defined as "*data that has been generated by computer algorithms or statistical models and does not contain personal data*", therefore considering synthetic data as de-identified data. The AI legislation in Utah is followed by Colorado [18], which will be effective on February 1, 2026. This legislation majorly focuses on the risks of AI systems related to discrimination, including many other arguments. "**Voice**" or "**synthetic data**" is not defined in the legislation, but this legislation does not consider conversational AI technology a high risk unless it generates content that is discriminatory or harmful. AI legislation has passed in only these two U.S. states, and in progress in several states: California, Illinois, New York, Louisiana, Massachusetts, Ohio, and Oklahoma.

The proposed California AI Transparency Act [19] would require any person that creates, codes, or otherwise produces a generative AI system, which is publicly available and has over 1,000,000 monthly visitors, to include a latent disclosure (permanent, to the extent it is technically feasible) in AI-generated digital content (synthetic data) including "audio". Illinois [20] Consumer Fraud-AI legislation proposes for requirement of disclosure on synthetic media in advertising. However, the definition also includes "human voice" *created, reproduced, or modified by generative AI or a software algorithm to produce or reproduce a human voice*. Similar to Illinois, the AI legislation in New York [21] requires advertisements to disclose the use of a synthetic performer (synthetic data), but does not explicitly mention "voice". Louisiana's proposed AI bill [22] primarily revolves around guidelines for AI foundation models, requiring every publicly available (made available by any person in the state) foundation model and its use to be officially registered with the state, and await for further guidelines from the state. "Audio" is explicitly defined as one of the "AI-generated content" (synthetic data) either created or modified by a "generative artificial intelligence system" in the proposed AI legislation for the Commonwealth of Massachusetts [23]. A mandatory disclosure is required for all generative AI systems, otherwise punishable, with notice and metadata information of the AI-generated content. Any person using such a Generative-AI system to generate or re-purpose AI-generated content would also be prohibited from removing the disclosure information. The proposed AI legislation in Ohio [24] provides guidelines on AI-generated (or synthetic) products and the prohibition of identity fraud using a replica of a person. "Replica of a person's persona" (or replica) is defined as a customized version of an individual's "voice" (and other factors), that appears to be the individual's authentic persona. The replica could be partially or fully generated by AI. The proposed AI legislation in Oklahoma [25] gives its citizens the right to consent to any "*derivative media that is generated by an artificial intelligence engine and uses audio recordings of the citizen's voice or images of him or her to recreate the citizen's likeness*".

6. Conclusion

Innovation in AI is necessary, but not at the cost of privacy. The speech technology community has led several efforts on voice privacy including The Voice Privacy Challenge [26] and Symposium on Security and Privacy in Speech Communication. It is important to communicate such efforts on voice privacy to the public and policymakers. Businesses have an obligation to disseminate whether they follow privacy-preserving technology development, and also how they collect, use, and maintain consumer data. Such actions would be valuable for speech technology research and development to sustain, as the world navigates regulations on privacy and trustworthy AI. On the legislative actions, momentum on both privacy and AI bills is rapid. Apart from many regulations, California and Illinois have proposed Consumer Privacy Funds to educate the public, including children in the area of online privacy. In the future, it would be interesting to navigate and compare how the legislative landscape in the U.S. differs from other countries across the globe such as the GDPR and the EU AI Act.

7. Acknowledgements

This work is sponsored by the Quad Fellowship and McDermott Graduate Fellowship (Dutta), and NSF awards #1918032, #2234916, and #2341384 (Hansen). Partial work was done for the URA Science Policy Competition (SPARC) and an AI Legislation Sprint by the Federation of American Scientists.

[1] Information captured and converted to a mathematical representation, including a numeric string or similar configuration, that cannot be used to recreate data generated by automatic measurement of an individual's biological patterns or characteristics used to identify the specific individual

8. References

[1] "The blueprint for an ai bill of rights." [Online]. Available: https://www.whitehouse.gov/ostp/ai-bill-of-rights/

[2] "President biden issues executive order on safe, secure, and trustworthy artificial intelligence." [Online]. Available: https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development.-and-use-of-artificial-intelligence/

[3] "Driving u.s. innovation in artificial intelligence: A roadmap for artificial intelligence policy in the united states senate." [Online]. Available: https://www.young.senate.gov/wp-content/uploads/Roadmap_Electronic1.32pm.pdf

[4] S. Dutta, S. A. Tao, J. C. Reyna, R. E. Hacker, D. W. Irvin, J. F. Buzhardt, and J. H. Hansen, "Challenges remain in Building ASR for Spontaneous Preschool Children Speech in Naturalistic Educational Environments," in *Proc. Interspeech 2022*, 2022, pp. 4322–4326.

[5] S. Dutta, D. Irvin, J. Buzhardt, and J. H. Hansen, "Activity focused speech recognition of preschool children in early childhood classrooms," in *Proceedings of the 17th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2022)*, 2022, pp. 92–100.

[6] S. Dutta, I. López-Espejo, D. Irvin, and J. H. L. Hansen, "Joint Language and Speaker Classification in Naturalistic Bilingual Adult-Toddler Interactions," in *Proc. The Speaker and Language Recognition Workshop (Odyssey 2024)*, 2024, pp. 81–85.

[7] "H.r.8818 - american privacy rights act of 2024." [Online]. Available: https://www.congress.gov/bill/118th-congress/house-bill/8818

[8] "Innovation, data, and commerce subcommittee markup recap: Monumental step forward for data privacy and kids online safety," the House Energy and Commerce Committee, May 23, 2024.

[9] "Children's online privacy protection act of 1998." [Online]. Available: https://www.ecfr.gov/current/title-16/chapter-I/subchapter-C/part-312

[10] "H.r.7890 - children and teens' online privacy protection act." [Online]. Available: https://www.congress.gov/bill/118th-congress/house-bill/7890

[11] "California consumer privacy act (ccpa)." [Online]. Available: https://oag.ca.gov/privacy/ccpa

[12] "Privacy rights act." [Online]. Available: https://ilga.gov/legislation/BillStatus.asp?GA=103&SessionID=112&DocTypeID=SB&DocNum=3517

[13] "Delaware personal data privacy act." [Online]. Available: https://legis.delaware.gov/BillDetail?LegislationId=140388

[14] "Consumer data privacy act." [Online]. Available: https://www.legis.state.pa.us/cfdocs/billinfo/bill_history.cfm?syear=2023&sind=0&body=H&type=B&bn=1201

[15] "Ohio personal privacy act." [Online]. Available: https://www.legislature.ohio.gov/legislation/135/hb345

[16] "Utah s.b. 149 - artificial intelligence amendments." [Online]. Available: https://le.utah.gov/%7E2024/bills/static/SB0149.html

[17] "Utah consumer privacy act." [Online]. Available: https://le.utah.gov/~2022/bills/static/SB0227.html

[18] "Colorado s.b. 205 - consumer protections for artificial intelligence." [Online]. Available: http://www.oklegislature.gov/BillInfo.aspx?Bill=hb3453&Session=2400

[19] "Sb-942 california ai transparency act." [Online]. Available: https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202320240SB942

[20] "Illinois h.b. 4869 - consumer fraud-ai disclosure." [Online]. Available: https://ilga.gov/legislation/BillStatus.asp?GA=103&SessionID=112&DocTypeID=HB&DocNum=4869

[21] "New york senate bill s6859a." [Online]. Available: https://www.nysenate.gov/legislation/bills/2023/S6859/amendment/A

[22] "Louisiana senate bill 118." [Online]. Available: https://www.legis.la.gov/Legis/BillInfo.aspx?s=24RS&b=SB118

[23] "Massachusetts h.b. 4788 - an act relative to artificial intelligence disclosure." [Online]. Available: https://malegislature.gov/Bills/193/HD4788

[24] "Ohio senate bill 217." [Online]. Available: https://www.legislature.ohio.gov/legislation/135/sb217

[25] "Oklahoma house bill 3453." [Online]. Available: http://www.oklegislature.gov/BillInfo.aspx?Bill=hb3453&Session=2400

[26] N. Tomashenko, X. Wang, E. Vincent, J. Patino, B. M. L. Srivastava, P.-G. Noé, A. Nautsch, N. Evans, J. Yamagishi, B. O'Brien *et al.*, "The voiceprivacy 2020 challenge: Results and findings," *Computer Speech & Language*, vol. 74, p. 101362, 2022.

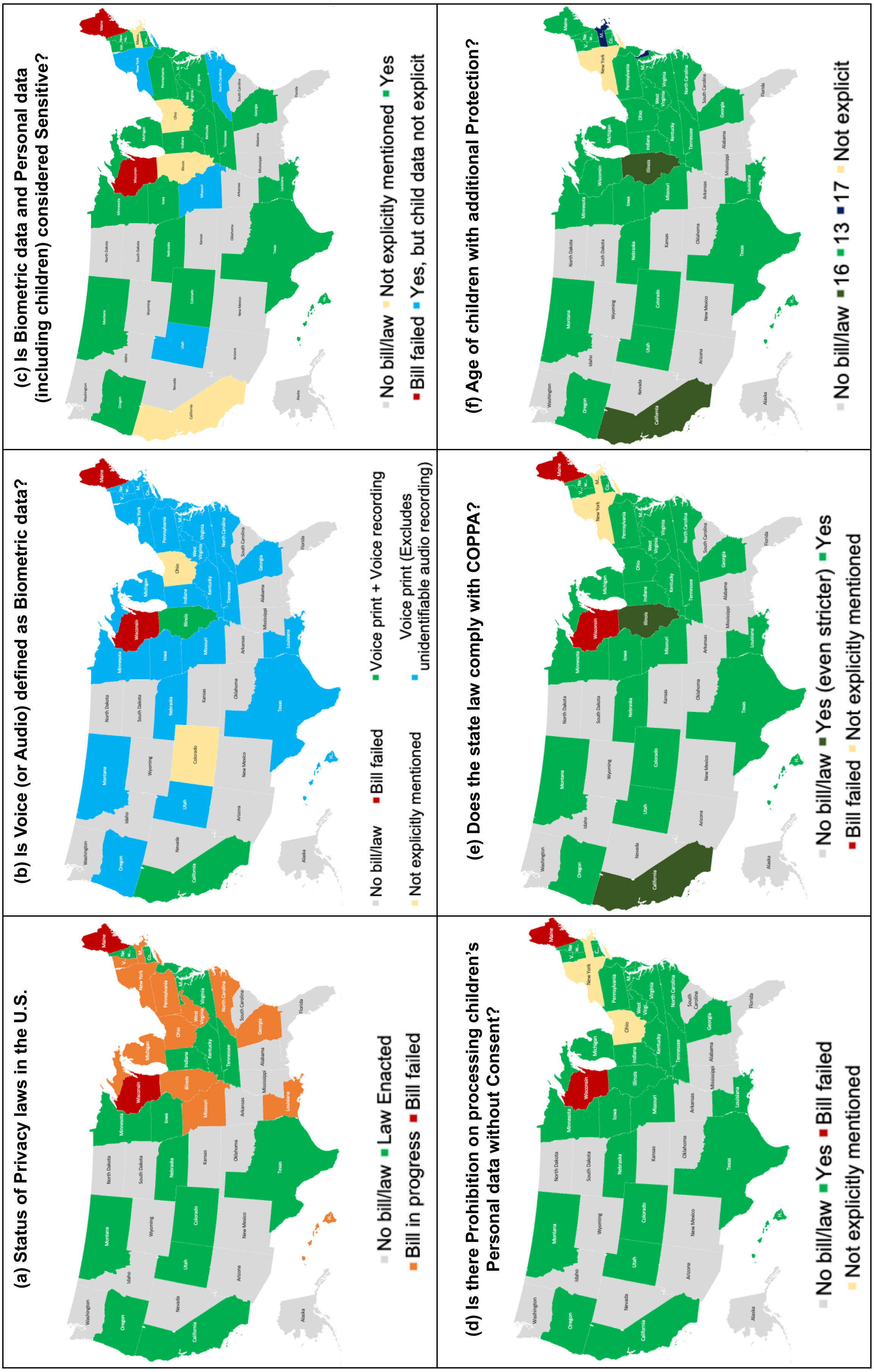
(a) Status of Privacy laws in the U.S.
⬜ No bill/law 🟩 Law Enacted
🟧 Bill in progress 🟥 Bill failed
(b) Is Voice (or Audio) defined as Biometric data?
⬜ No bill/law 🟥 Bill failed 🟩 Voice print + Voice recording
🟨 Not explicitly mentioned 🟦 Voice print (Excludes unidentifiable audio recording)
(c) Is Biometric data and Personal data (including children) considered Sensitive?
⬜ No bill/law 🟨 Not explicitly mentioned 🟩 Yes
🟥 Bill failed 🟦 Yes, but child data not explicit
(d) Is there Prohibition on processing children's Personal data without Consent?
⬜ No bill/law 🟩 Yes 🟥 Bill failed
🟨 Not explicitly mentioned
(e) Does the state law comply with COPPA?
⬜ No bill/law 🟩 Yes (even stricter) 🟩 Yes
🟥 Bill failed 🟨 Not explicitly mentioned
(f) Age of children with additional Protection?
⬜ No bill/law 🟩 16 🟩 13 🟦 17 🟨 Not explicit

Figure 2: *(a) Status of Privacy laws in the U.S. (b) Is Voice (or Audio) defined as Biometric data? (c) Is Biometric data and Personal data (including children) considered Sensitive? (d) Is there a Prohibition on processing children's Personal data without Consent? (e) Does the state law comply with COPPA? (f) Age of children with additional Protection?*