Data Protection, Privacy and US Regulation

Denise DiPersio
Linguistic Data Consortium, University of Pennsylvania
3600 Market Street, Suite 810, Philadelphia, PA 19104 USA
dipersio@ldc.upenn.edu

Abstract

This paper examines the state of data protection and privacy in the United States. There is no comprehensive federal data protection or data privacy law despite bipartisan and popular support. There are several data protection bills pending in the 2022 session of the US Congress, five of which are examined in Section 2 below. Although it is not likely that any will be enacted, the growing number reflects the concerns of citizens and lawmakers about the power of big data. Recent actions against data abuses, including data breaches, litigation and settlements, are reviewed in Section 3 of this paper. These reflect the real harm caused when personal data is misused. Section 4 contains a brief US copyright law update on the fair use exemption, highlighting a recent court decision and indications of a re-thinking of the fair use analysis. In Section 5, some observations are made on the role of privacy in data protection regulation. It is argued that privacy should be considered from the start of the data collection and technology development process. Enhanced awareness of ethical issues, including privacy, through university-level data science programs will also lay the groundwork for best practices throughout the data and development cycles.

Keywords: data protection, privacy, regulation

1. Introduction[1]

This is an interesting time for the field of language resources and related technologies. From the first days of natural language processing research represented by early machine translation, document understanding and speech recognition systems, we are today surrounded by human language technologies that are part of our daily lives. How we got here is a story about lots of good people doing good work in academia and industry and not least, sharing data broadly among the community. Data sharing has been fraught with legal issues, principally copyright rights and related licensing considerations, and depending on the data type, ethics and privacy concerns. Some of those issues persist in commercial language technologies, affecting how the systems work and how an individual's data is protected. The work grew faster than the law, so we find ourselves trying to match law and ethics with today's research and business realities. Tensions abound.

This paper examines the state of data protection and privacy in the United States, where the catching-up process has a long way to go. There is still no comprehensive federal data protection or data privacy law that addresses key issues. In the meantime, several US states have passed laws of their own (and more are in the works), and some federal agencies, principally the US Federal Trade Commission, investigate data-related consumer harms. The lack of an overarching philosophy or schema is a real problem.

There are several data protection bills pending in the 2022 session of the US Congress, five of which are examined in Section 2 below. Although it is not likely that any will be enacted, the growing number reflects the concerns of citizens and lawmakers about the power of big data.

Recent actions against data abuses, including data breaches, litigation and settlements, are reviewed in Section 3 of this paper. These reflect the real harm caused when personal data is misused.

Section 4 contains a brief US copyright law update on the fair use exemption, highlighting a recent court decision and indications of a re-thinking of the fair use analysis.

In Section 5, some observations are made on the role of privacy in data protection regulation. It is argued that privacy should be considered from the start of the data collection and technology development process. Enhanced awareness of ethical issues, including privacy, through university-level data science programs will also lay the groundwork for best practices throughout the data and development cycles.

2. Data Protection

2.1 Lack of US Progress

As reported at LREC2018, there is no comprehensive data protection law in the United States, and those that exist apply mostly to government use of personal information or to special circumstances (e.g., health information, personal credit information, student education records, children's online activity). (DiPersio, 2018). Private organizations face little regulation with respect to the collection, storage and use of data collected from or about individuals in the course of their business. This cuts across all industries, but is especially problematic with respect to the large technology companies that dominate the US, and to some extent, the global, economy.

Enacting a comprehensive US data protection scheme is an issue that has some level of bipartisan political support in Congress as well as broad popular appeal, but little progress has been made to date. The situation is becoming urgent, however, as individuals become increasingly aware of the ways in which their personal information is being used (and exploited) in the digital space. Companies claim to self-regulate, but those efforts often fall short. Several states have their own data protection statutes, but standards and provisions vary. Victims of data breaches and other unfair or deceptive data practices can resort to the courts

[1]This paper does not provide legal advice and nothing in this paper should be construed to constitute legal advice.

and to some government agencies under various theories and laws, with the attendant possibility of inconsistent outcomes.

In an age of virtual, cross-border data flows, this data protection gap also affects US relations with other countries, a growing number of which, led by the European Union and the GDPR, have enacted comprehensive data privacy laws. Indeed, some believe that the effect of the *Schrems II* decision, in which the European Court of Justice found that US data surveillance laws did not pass muster under the GDPR, could be ameliorated to some extent by US laws mandating standards for companies' collection and storage of personal information, thus in turn, limiting the reach of the US government's access to such information.

2.2 Pending Data Protection/Privacy Legislation

Several bills around the privacy and protection of an individual's personal data are pending in the 2021-2022 session of the US Congress. These include proposals that were introduced in the previous Congressional session (2019-2020), were not acted upon and were re-introduced in the current session. Most commentators believe that it is unlikely that any will be considered or enacted in this session, absent a showing of strong will from Congress and the Executive Branch.

Five of these bills are described below. Of these, three were introduced by members of the Democratic party (President Biden's party) (D), and two were introduced by members of the Republican party (R). Only one has co-sponsors from both parties. Four were pending in the 2019-2020 session and were re-introduced in 2021; no action (hearings, debates, etc.) has been taken on any of these bills in 2022 as of this writing.

These schemes represent varying approaches. Some are more comprehensive than others, some would preempt state data protection/privacy laws, some create new government agencies, and some rely on existing government infrastructure for investigation and enforcement. The more comprehensive proposals have some exemption for research-related activities. In all cases, existing federal data privacy/protection laws would remain in effect.

2.2.1 Information Transparency & Personal Data Control Act (D)

The **Information Transparency & Personal Data Control Act** (H.R. 1816) was introduced into the US House of Representatives in March 2021. This proposed law focuses on strengthening the powers of the US Federal Trade Commission (FTC), the agency with the authority to investigate unfair trade practices and to date the leading US regulator to take action on complaints alleging data abuses. It would provide the FTC with the authority to regulate the collection, processing, use, and storage of "sensitive personal information," an inclusive category that covers categories like financial account numbers, usernames and passwords, genetic data, citizenship, gender identity, web browsing history and more. Organizations that collect, store, process, sell, share or otherwise use sensitive personal information from more than 250,000 people annually would be required to undergo a privacy audit every two years. The act's restrictions do not apply to activities in the "public interest" including research, as long as processing does not create "significant harm" to users. [2]

This bill would also preempt state privacy laws. H.R. 1816 is considered to be more business-friendly than other proposals.

2.2.2 Data Protection Act (D)

In June 2021, the **Data Protection Act of 2021** (S. 2134) was introduced in the US Senate (S. 2134). (A similar bill was introduced in 2020, but no action was taken before the 2019-2020 Congressional session ended.) It provides for the creation of a federal Data Protection Agency that would be charged with developing and enforcing data protection rules. It includes sections around agency authority to review mergers involving large technology companies, or any merger that involves the transfer of the personal data from more than 50,000 individuals; the establishment of an Office of Civil Rights; and the ability to impose fines and punitive penalties for unlawful, unfair, deceptive, abusive or discriminatory data practices.

The bill focuses on "data aggregators" and "high risk data practices," both of which may require some further clarification regarding research-related uses. A data aggregator is defined as any person collecting, using or sharing personal data that is not "de minimis," exempting individuals who collect, user or share such data for non-commercial purposes.[3] "High risk data practices" include "a systematic processing of publicly accessible data on a large scale." [4]

Although the term "personal identifying information" is used throughout the US research community, including by US government agencies, it has no conclusive definintion. The Data Protection Act takes a broad approach, defining "personal data" as electronic data that "identifies, relates to, describes, is capable of being associated with, or could reasonably be linked, directly or indirectly, with a particular individual, household or device." [5]

There are no provisions addressing research-related exemptions, except perhaps implicitly by the above reference to those who collect and share data for non-commercial purposes. State laws offering greater protections than those under this act would not be preempted (e.g., California's data privacy law).

2.2.3 Filter Bubble Transparency Act (R, D co-sponsors)

The Filter Bubble Transparency Act, previously incorporated in the 2019 version of the SAFE DATA Act (see below), was re-introduced in June 2021 as separate legislation in the Senate (S. 2024). This act would require internet platforms to provide users with "the option to engage with a platform without being manipulated by

[2]H.R. 1816, Section 3(b)(1)(G).

[3]S. 2134, Section 2 (6).

[4]Ibid., Section 2 (11).

[5]Ibid., Section 2 (16). Compare to GDPR Article 4 (1) where "personal data" can refer to name, identification number, location, online presence, or physical, genetic, economic, cultural or social identity.

algorithms driven by user-specific data."[6] Specifically, users would have the option of a "filter bubble-free view" of information, the presentation or order of which is not determined by an "opaque" algorithm.

A platform conducting not-for-profit research is exempt from the bill.[7] The FTC would enforce violations of the act under its jurisdiction to investigate unfair or deceptive acts or practices. [8]

Google has publicly expressed concern about the act, telling its business users that it "could disrupt many of the digital tools you use" and "[m]ake it harder for customers to find you."[9]

2.2.4 Setting an American Framework to Ensure Data Access, Transparency and Accountability Act (SAFE DATA Act) (R)

Another bill from the 2019-2020 Congressional session reintroduced in July 2021 is the SAFE DATA Act (S. 2499). This Senate bill aims to give users control over how their data is accessed, used and maintained, to require businesses to follow transparent data practices, and to strengthen the FTC's rulemaking ability and enforcement authority. The legislation would preempt state privacy laws.

Of interest here is the definition of research. Processing data for a research purpose means that the "advancement of scientific knowledge" is the primary purpose of the activity, but it can also be for the commercial benefit of the entity processing the data.[10] Exempt from the bill are data collection, processing, and related activities conducted for research (peer-reviewed, public, historical, statistical) that follow applicable privacy and ethical laws including Institutional Review Board review under the federal regulations for human subjects research.[11]

2.2.5 Online Privacy Act (D)

Like most of the current propsals, the **Online Privacy Act** was pending in the 2019-2020 Congressional session but failed to advance. In November 2021, the bill (H.R. 6027) was reintroduced in the US House of Representatives. It gives users the right to access, correct or delete their data, limits the amount of data companies can collect, allows users to decide how long companies can maintain their data, and requires that companies obtain consent from users. A new Data Privacy Agency would be responsible for enforcement and investigation. Qualified research entities conducting work for non-commercial purposes would be exempt from the act's ban on re-identifying de-identified data.[12]

2.3 The Pitfalls of Lagging Behind

The divergent approaches to US data protection legislation illustrated in the selected bills above suggest that finding common ground will be a challenging task. In addition to an extremely partisan congressional atmosphere, other high priority issues such as infrastructure, the Ukraine war, climate change and more vie for lawmakers' attention. The likelihood that a data protection law will be enacted before the current session ends in January 2023 is doubtful. Nevertheless, at a Global Privacy Summit in April 2022, Congressional aides indicated that talks have been ongoing behind the scenes and as a result, some compromises are possible. The main points of contention are federal preemption (the continued viability of state privacy laws) and whether individuals/groups can sue companies for money damages under a federal law. A compromise that allows some state law provisions to remain and that permits a limited right of private action is apparently gaining traction. [13] But the timing of any solution is still unclear.

The recent settlement proposal in the Clearview AI litigation, a case brought by the ACLU and others based in part on Illinois' biometric data statute (which requires user consent to the use of biometric data, including faces), reveals the shortcomings when state laws must fill the data protection gap. (See Section 3.2 below). The settlement will have some broad applicability to the extent that Clearview will not be able to sell its faces database to most US companies, but only photographs taken in, or uploaded from, Illinois will be removed from the database. The lack of a federal law regulating personal biometric information means in this case a less than satisfactory result

A larger issue, however, is that the United States is out of step with the global community in its piecemeal approach, shunning a data protection law that cuts across specific use cases. This is not new; traditional American thinking regards regulation as an impediment to innovation and US competitive standing. The GDPR, on the other hand, reflects broad goals regarding fundamental rights and economic and social issues.[14] Indeed the GDPR is viewed as setting the international standard for data protection and privacy. Other countries are moving forward with their own data protection and privacy regimes, many of which are based on, or are similar to, the GDPR model.[15]

---

[6]S. 2024, Preamble.

[7]Ibid., Section 2(4)(B)(III)(ii).

[8]Ibid., Section 4(a).

[9]Boyle, Christopher. *Google Fear of Looming "Filter Bubble Transparency Act" Legislation Which Would Force Fairness, Disclosure and Accountability*. Available at : https://www.publishedreporter.com/2021/11/24/google-scared-of-looming-filter-bubble-transparency-act-which-would-force-fairness-disclosure-and-accountability/.

[10]S. 2499, Section 2(16).

[11]Ibid., Section 108(a)(10).

[12]H.R. 6027, Section 205(c).

[13]Lima, Cristiano. The debate over a privacy bill is inching forward on Capitol Hill. Available at: https://www.washingtonpost.com/politics/2022/04/13/debate-over-privacy-bill-is-inching-forward-capitol-hill/.

[14]Roberts, Huw and Luciano Floridi. The EU and the US: two different approaches to AI governance. Available at: https://venturebeat.com/2022/03/21/why-2022-is-only-the-beginning-for-ai-regulation/.

[15]Those include the United Kingdom, Switzerland, Turkey, Australia, China, India, Indonesia, Japan, New Zealand, Philippines, Singapore, Thailand, South Africa, Saudi Arabia and

The 2020 decision of the European Court of Justice in *Schrems II* that US privacy safeguards were not "adequate" within the meaning of the GDPR is one example of how the philosophy gap between the United States and other countries affects international commerce.[16] The European court was concerned specifically about US intelligence laws that allow broad access to individual data.[17] In March 2022, the US and EU reached a tentative agreement on the dispute, with the US agreeing to make some administrative adjustments to intelligence law procedure that the parties believe will support a US claim that US safeguards are "necessary and proportionate in the pursuit of defined national security objectives." [18] However, because US law will not be changed, many believe that this latest agreement will be challenged in court as well. Again, the US line, here its stated need for surveillance, is at odds with the EU's focus on personal data protection.[19]

3. Data Breaches, Litigation, Settlements

Data breaches caused by events like cyberattacks, human errors, malicious activity and negligence are a daily threat for anyone whose personal information is stored in some organization's database. The full extent of the damage caused by US data breaches is often hard to assess since there are few regulations requiring that breaches be reported and the scope of any damage revealed. Too often, users discover that their personal information was compromised long after the event. The US Privacy Rights Clearinghouse is a non-profit organization with the goal to protect privacy for all. As part of that work, it has tracked reported US data breaches since 2005. It currently reports that over 11 billion records were compromised in more than 9000 reported US data breaches since 2005.[20] The actual numbers are likely much higher.

In the meantime, victims of data breaches or data misuse have been calling companies to account. Below are some recent examples.

3.1 Data Abuse Victims React

Selected challenges to US data breaches and related violations in 2021-2022 include the following.

Online merchandise platform **CafePress** settled FTC claims that it failed to secure users' sensitive personal data and failed to disclose a major data breach that allowed hackers to access millions of email addresses, passwords, social security numbers, credit card information and more. The company's former owner will pay $500,000 to data breach victims and along with the current owner, will implement security measures to address the circumstances leading to the data breach. Financial services company **Plaid, Inc.** will pay $58 million to settle a legal action in which Plaid was accused of accessing personal banking information without consent from users of financial applications such as Robinhood and Venmo. **Zoom** agreed to an $85 million settlement in a California federal lawsuit alleging that it engaged in unauthorized sharing of user data, misrepresented its encryption services and allowed hackers to disrupt meetings. **OpenX Technologies**, an advertising platform, must pay $2 million to settle claims by the FTC that it collected data from children in violation of the agency's Children's Online Privacy Protection Act Rule. **TikTok** settled consolidated litigation alleging that it shared users' personal data without consent, improperly handled users' biometric data and engaged in ad targeting for $92 million; the case involved roughly 89 million users. **Meta/Facebook** agreed to settle two privacy class-action lawsuits: it will pay $650 million to resolve allegations that it tagged biometric information in violation of Illinois law (2020) and $90 million to settle claims made in 2012 that it tracked users' activity after they logged off the platform (2022).

3.2 Clearview AI Litigation

In 2020, US facial recognition company Clearview AI claimed that it had developed a database of three million human images scraped from the web and annotated for biometric characteristics which it made available to law enforcement organizations and other paying customers. That generated a series of lawsuits from affected groups alleging breach of privacy and related theories. The cases were consolidated in a Chicago, Illinois federal court; they include claims under Virginia, California and New York law and under Illinois' Biometric Information Privacy Act (BIPA). BIPA constrains how companies can collect, use and store biometric information and requires that such information cannot be collected or used without users' written consent. It also permits individuals to bring actions under the statute on their own behalf.

Brazil. Gibson Dunn. International Cybersecurity and Data Privacy Outlook and Review – 2022. Available at: https://www.gibsondunn.com/international-cybersecurity-and-data-privacy-outlook-and-review-2022/.

[16]The so-called "privacy shield" relates to cross-border data transfers of personal information.

[17]Schaetzel, Lucas, J. U.S. and E.U. Reach New Trans-Atlantic Data Flow Agreement To Replace Privacy Shield. Available at: https://www.beneschlaw.com/resources/us-and-eu-reach-new-trans-atlantic-data-flow-agreement-to-replace-privacy-shield.html.

[18]FACT SHEET: United States and European Commission Announce Transatlantic Data Privacy Framework. Available at: https://www.whitehouse.gov/briefing-room/statements-releases/2022/03/25/fact-sheet-united-states-and-european-commission-announce-trans-atlantic-data-privacy-framework/.

[19]Ikeda, Scott. EU and US Move Closer to Privacy Shield Replacement With Agreement on Data Transfer Deal. Available at: https://www.cpomagazine.com/data-privacy/eu-and-us-move-closer-to-privacy-shield-replacement-with-agreement-on-data-transfer-deal/.

[20]Privacy Rights Clearinghouse. Available at: https://privacyrights.org/.

YouTube, Twitter and others demanded that Clearview stop collecting images from their sites in early 2020, claiming that the company's acts violated the sites' terms of use.

In the meantime, the company continued to conduct business. It announced that it was collecting 100 billion photos for its database, processing around 1.5 billion images monthly.

The presiding judge in the Illinois case ruled against Clearview's motion to dismiss the complaint in February 2022. In May 2022, a proposed settlement to the litigation was announced under which Clearview agreed to change its business model. Specifically, it will only sell its algorithm to customers, not its faces database. Additionally, it will not work with Illinois police or government organizations for five years and will remove from its database all photos taken in, or uploaded from, Illinois.[21]

This result shows the value of a statute like BIPA. Clearview's business practices are also under review in various foreign venues.[22]

3.3 Data Brokers, Cloud Services, Cyberattacks

A recent lawsuit involving two US **data brokers** illustrates the downstream risks associated with the collection of individuals' personal data. The litigation involves Outlogic (formerly known as X-Mode) and NybSys. X-Mode sells location data it collects from various apps, and it licensed that data to NybSys. X-Mode claims that NybSys unlawfully resold the information to another data broker, that in turn resold it to others. The suit is based on contract and trade secret claims. The case was referred to private mediation in March 2022.

Cloud services have become vital to everyday life, comparable in some ways to essential business such as power companies. Yet, those services are controlled by a few actors – Alphabet, Amazon, Microsoft – that are essentially unregulated. Missing are obligations around reporting data breaches. A model for oversight could be a recently-enacted US law requiring "key businesses" to report **cyberattacks/hacks** within 72 hours to the government's Cybersecurity and Infrastructure Security Agency, part of the US Department of Homeland Security. Covered businesses include banks and utilities. Any ransomware payments must be reported within 24 hours of the payment. Details on coverage, reporting, deadlines and so on are to be worked out in forthcoming regulations. Industry groups have criticized the measure as likely to result in large amounts of non-meaningful information that will hinder government analysis; they urge time for companies to assess the extent of any breach and to assemble information targeting actual harm.

4. US Copyright Update

The trend in US courts has been to recognize that copyrighted materials used for machine learning purposes are eligible for the US copyright law's fair use exception. Principally based on the transformative nature of the machine learning use case, such rulings have included unmodified full-text searchable databases within the exception, as discussed in previous LREC workshops (DiPersio, 2018).

An interesting 2021 case involving photographs of the musical artist Prince raised the relationship between derivative works under US copyright law (in which the original rightsholder shares copyright with the derivative works author) and fair use. Finding that "[i]t does not follow . . . that any secondary work that adds a new aesthetic or new expression to its source material is necessarily transformative," a US federal court held that changes made by Andy Warhol to the Prince photographs without the original author's permission were "substanially similar" to the orignal works and therefore more derivative than transformative.[23] Although this ruling may be deemed applicable to artistic works only, it bears watching to the extent courts could be influenced to take a harder look at transformation generally.

This ties in with the view of some that the emphasis on transformativeness in the fair use analysis overlooks the traditional thinking that fair use is supposed to benefit the less powerful non-rights holder against the monopoly of the copyright holder. Instead, as one scholar claims, "[t]oday's tech business turns this structure on its head," allowing "big users" to monetize lots of "little content" that includes allowing machines to learn from the way authors express ideas.[24] Moreover, with respect to the fair use factor around whether a substitute market (e.g., for machine learning) exists for the original rightsholder, a question that courts have in the past answered in the negative, one can imagine that text owners today would take advantage of the existing market for training data, for example. The data science and machine learning communities have benefited from the fair use copyright exception, but the growing power of technology, its insatiable need for data and the demonstrated ways in which individuals are harmed by big

[21]Harwell, Drew. Clearview AI to stop selling facial recognition tool to private firms. Available at: https://www.washingtonpost.com/technology/2022/05/09/clearview-illinois-court-settlement/.

[22]France 24. Clearview AI agrees to limit sales of facial recognition data after ACLU lawsuit. Available at: https://www.france24.com/en/americas/20220510-clearview-ai-settles-suit-agrees-to-limit-sale-of-facial-recognition-database (referencing proceedings in Canada, Italy, France, Austria and the United Kingdom).

[23]The Andy Warhol Foundation for the Visual Arts, Inc. V. Lynn Goldsmith, Lynn Goldsmith, Ltd., 11 F.4th 26, 38, 42 (2d Cir. 2021). The US Supreme Court has agreed to hear an appeal of this decision in its fall 2022 term.

[24]Sobel, B. L. W. (2017). Artificial Intelligence's Fair Use Crisis. The Columbia Journal of Law & The Arts, 41(1), pp. 45–97, 87, 89. Available at : https://doi.org/10.7916/jla.v41i1.2036.

data may cause policymakers and courts to rethink their approach.

5. Privacy and Regulation

A final word about privacy. Even as the number of US bills to address technology-related data protection and personal privacy issues increase in number, some question using established legal principles to address the ways the digital world impacts personal information. Protecting persons from intrusion is rooted in the idea of a private space sacrosanct to the individual. Thus, the notions of "zones of privacy" and a person's "expectation of privacy" – developed in the late 19th century in connection with the inventions of the telephone and photography and applied to new, related technologies (e.g., wiretapping) through the 20th century and beyond (Chertoff, 2018) – were coined to describe such boundaries. A companion principle is that information provided "voluntarily" to third parties is not protected. The distinction between public and private information, however, is blurred in the digital space.

Privacy should therefore be considered less rigidly, as something that applies variously depending on the information and the context. (Hartzog, 2018). The suggestion has been made that *autonomy* or *human values* are better expressions of the notions underlying privacy because they take into account the mass of personal information collected, processed, repurposed and resold today. (Ibid.; Chertoff, 2018). Moreover, to be effective, such personal human values should be considered at the beginning of the development process, not after the technology or application is finished and operational, because assumptions that implicate privacy are incorporated at the start: "[w]e shape our tools and thereafter our tools shape us."[25] This can be as innocuous as including features for convenience or responding to corporate pressure to generate data so that it can be monetized downstream. Some of this behavior is occasionally explained as resulting in unintended consequences. Nevertheless, design choices are not necessarily value neutral; they can favor certain societal interests over others. (Ibid.).

The move to create or enhance data science programs in US colleges and universities offers an opportunity to make ethics, privacy and related issues part of the curriculum, and many institutions offer such courses and training. (Baumer, et al. 2022; Davis, 2020). It is also encouraging to note that some large companies, including Google, Apple and Facebook, have implemented internal processes for evaluating privacy and ethical issues in their data collection and research activities.[26] Behind the scenes, however, is the specter of artificial intelligence and its boundless capabilities. Companies are spending substantial sums on "AI" research.[27] Academic research, sometimes with industry partners, reflects this trend as well. Thus, the tension between the technology industry's continued need for more researchers (software engineers, linguists) to advance the corporate mission and the goal of developing technology that serves all interests of society.

6. Conclusion

This paper has attempted to unite several themes around the regulation of data protection and privacy in the United States: the state of federal legislative initiatives, legal proceedings relating to data abuses, and thinking about how traditional notions about privacy relate or not to the realities of digital life. As discussed above, the current failure of a principled approach to regulating data protection and privacy in the United States means that those most vulnerable – everyone whose data is collected, analyzed, shared and sold – have little clarity on achieving effective relief. Pending legislation addresses some aspects of the problem, but until federal laws are enacted, data abuse victims must resort to the courts and to administrative remedies under a variety of legal frameworks. Those developing the means to exploit, or those exploiting user data, are the beneficiaries for now, although recent legal decisions and settlements suggest that the situation may be changing. Nevertheless, the United States lags behind its international partners in dealing with the digital world. The *Andy Warhol Foundation* copyright case illustrates another prospect for change, namely a re-thinking of how the transformation test could be applied in future cases involving machine learning applications. Similarly, a new notion of privacy that abandons the traditional US legal concept could lead to more effective regulation with respect to personal information in the digital space. And providing students and the professional community with ethics training and better tools for navigating the research and development process has the prospect of mitigating the occurrence of unintended consequences.

7. Bibliographical References

Altman, Micha, et al. (2018). Practical approaches to big data privacy over time. International Data Privacy Law, 8 (1), pp. 29-51.

Associated Press. *Facebook, YouTube demand facial recognition company stop scraping faces from sites*. Available at: https://www.nbcnews.com/tech/security/facebook-youtube-demand-facial-recognition-company-stop-scraping-faces-sites-n1131786.

Baumer, Benjamin S. et al. (2022). Integrating data science ethics into an undergraduate major; A case study. Available at https://arxiv.org/pdf/2001.07649.pdf.

Boyle, Christopher. *Google Fear of Looming "Filter Bubble Transparency Act" Legislation Which Would Force Fairness, Disclosure and Accountability*.

[25]Hartzog, Woodrow. (2018). *Privacy's Blueprint,* 8 n.11. Cambridge, Massachusetts: Harvard University Press (quoting Marshall McLuhan).

[26]Altman, Micha, et al. (2018). Practical approaches to big data privacy over time. International Data Privacy Law, 8 (1), pp. 29-51, 38.

[27]Rosenbush, Steven. *Big Tech Is Spending Billions on AI Research. Investors Should Keep An Eye Out.* Available at: https://www.wsj.com/articles/big-tech-is-spending-billions-on-ai-research-investors-should-keep-an-eye-out-11646740800.

Available at : https://www.publishedreporter.com/2021/11/24/google-scared-of-looming-filter-bubble-transparency-act-which-would-force-fairness-disclosure-and-accountability/.

Channick, Robert. *Nearly 1.6 million Illinois Facebook users could get their $400 checks soff after appeals court upholds $650 million settlement.* Available at: https://www.chicagotribune.com/business/ct-biz-facebook-privacy-settlement-illinois-appeal-decision-20220317-e4s3jqzm7bfvxliwh2uibbfo7y-story.html.

Chertoff, M. (2018). *Exploding Data, Reclaiming Our Cybersecurity In The Digital Age*. New York, New York: Atlantic Monthly Press.

Davis, Karen C. (2020). Ethics in Data Science Education. *In American Society for Engineering Education Virtual Conference.* Available at: https://peer.asee.org/ethics-in-data-science-education.pdf.

Dean, Graham and Ronald Raether. *U.S. Senators Reintroduce Privacy Legislation*. Available at: https://www.jdsupra.com/legalnews/u-s-senators-reintroduce-privacy-4527842/#_ftn6.

DiPersio, D. (2018). *A US Perspective on Selected Legal and Ethical Issues Affecting the Development of Language Resources and Related Technology.* In Nicoletta Calzolari (Conference Chair), et al., editors, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC'18), W21, Legal Issues and Ethics*, Miyazaki, Japan, May. European Language Resource Association (ELRA).

EU General Data Protection Regulation (GDPR): 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), OJ 2016 L 119/1.

FACT SHEET: United States and European Commission Announce Transatlantic Data Privacy Framework. Available at: https://www.whitehouse.gov/briefing-room/statements-releases/2022/03/25/fact-sheet-united-states-and-european-commission-announce-trans-atlantic-data-privacy-framework/.

Federal Trade Commission. *Advertising Platform OpenX Will Pay $2 Million for Collecting Personal Information from Children in Violation of Children's Privacy Law*. Available at: https://www.ftc.gov/news-events/news/press-releases/2021/12/advertising-platform-openx-will-pay-2-million-collecting-personal-information-children-violation.

Federal Trade Commission. *FTC Takes Action Against CafePress for Data Breach Cover Up.* Available at: https://www.ftc.gov/news-events/news/press-releases/2022/03/ftc-takes-action-against-cafepress-data-breach-cover.

France 24. *Clearview AI agrees to limit sales of facial recognition data after ACLU lawsuit.* Available at: https://www.france24.com/en/americas/20220510-clearview-ai-settles-suit-agrees-to-limit-sale-of-facial-recognition-database.

Gibson Dunn. *International Cybersecurity and Data Privacy Outlook and Review – 2022.* Available at: https://www.gibsondunn.com/international-cybersecurity-and-data-privacy-outlook-and-review-2022/.

Gold, Ashley. *Exclusive: New bipartisan bill takes aim at algorithms.* Available at: https://www.axios.com/algorithm-bill-house-bipartisan-5293581e-430f-4ea1-8477-bd9adb63519c.html.

Haasch, Palmer. *TikTok may owe you money from its $92 million data privacy settlement*. Available at: https://www.businessinsider.com/tiktok-data-privacy-settlement-how-to-submit-claim-2021-11.

Hammer, Alex. *Facial Recognition firm Clearview AI says it will soon have 100 BILLION photos in its database to ensure 'almost everyone in the world will be identifiable' and wants to expand beyond law enforcement*. Available at: https://www.dailymail.co.uk/news/article-10523739/Clearview-AI-seeking-100-billion-photos-facial-recognition-database.html.

Hartzog, Woodrow. (2018). *Privacy's Blueprint*. Cambridge, Massachusetts: Harvard University Press.

Harwell, Drew. *Clearview AI to stop selling facial recognition tool to private firms.* Available at: https://www.washingtonpost.com/technology/2022/05/09/clearview-illinois-court-settlement/.

Holland, Makenzie. *Federal data privacy law efforts fizzle.* Available at: https://www.techtarget.com/searchcio/news/252512860/Federal-data-privacy-law-efforts-fizzle?vgnextfmt=print.

Ikeda, Scott. *EU and US Move Closer to Privacy Shield Replacement With Agreement on Data Transfer Deal. Available* at: https://www.cpomagazine.com/data-privacy/eu-and-us-move-closer-to-privacy-shield-replacement-with-agreement-on-data-transfer-deal/.

Jaehnig, Johnathan. *YouTube Claims to Be "Explicitly Clear" on Facial Recognition, But Is It Really?* Available at: https://www.makeuseof.com/youtube-claims-to-be-explicitly-clear-on-facial-recognition-but-is-it-really/.

Keegan J. and Alfred Ng. *Lawsuit Highlights How Little Control Brokers Have Over Location Data.* Available at: https://themarkup.org/privacy/2022/03/21/lawsuit-highlights-how-little-control-brokers-have-over-location-data.

Keegan J. and Alfred Ng. *There's a Multibillion-Dollar Market for Your Phone's Location Data.* Available at: https://themarkup.org/privacy/2021/09/30/theres-a-multibillion-dollar-market-for-your-phones-location-data.

Kerry, Cameron F. *One year after Schrems II, the world is still waiting for U.S. privacy legislation*. Available at: https://www.brookings.edu/blog/techtank/2021/08/16/one-year-after-schrems-ii-the-world-is-still-waiting-for-u-s-privacy-legislation/.

Lexis/Nexis. *AI in Academia: How the Need for Future Data Scientists & the Availability of Big Data is Transforming Universities.* Available at: https://www.lexisnexis.com/community/insights/professional/b/trends/posts/ai-in-academia.

Lima, Cristiano. *Europe is lapping the U.S. on tech regulation – again.* Available at: https://www.washingtonpost.com/politics/2022/03/28/europe-is-lapping-us-tech-regulation-again/.

Lima, Cristiano. *The debate over a privacy bill is inching forward on Capitol Hill.* Available at: https://www.washingtonpost.com/politics/2022/04/13/debate-over-privacy-bill-is-inching-forward-capitol-hill/.

Meltzer, Joshua P. *The Court of Justice of the European Union in Schrems II: The impact of GDPR on data flows and national security.* Available at: https://www.brookings.edu/research/the-court-of-justice-of-the-european-union-in-schrems-ii-the-impact-of-gdpr-on-data-flows-and-national-security/.

Michaels, Daniel and Sam Schechner. *U.S., EU Reach Preliminary Deal on Data Privacy.* Available at: https://www.wsj.com/articles/u-s-eu-reach-preliminary-deal-on-data-privacy-11648200085.

Privacy Rights Clearinghouse. Available at : https://privacyrights.org/

Roberts, Huw and Luciano Floridi. *The EU and the US: two different approaches to AI governance.* Available at: https://venturebeat.com/2022/03/21/why-2022-is-only-the-beginning-for-ai-regulation/.

Rosenbush, Steven. *Big Tech Is Spending Billions on AI Research. Investors Should Keep An Eye Out.* Available at: https://www.wsj.com/articles/big-tech-is-spending-billions-on-ai-research-investors-should-keep-an-eye-out-11646740800.

Roth, Emma. *Plaid, the service used by Venmo, Acorns, Robinhood, and more, may owe you some money.* Available at: https://www.theverge.com/2022/1/23/22898009/plaid-financial-venmo-acorns-robinhood-class-action-lawsuit.

Sobel, B. L. W. (2017). Artificial Intelligence's Fair Use Crisis. The Columbia Journal of Law & The Arts, 41(1), pp. 45-97. Available at; https://doi.org/10.7916/jla.v41i1.2036.

Spangler, Todd. *Meta to Pay $90 Million to Settle Decade-Old Facebook Data Privacy Lawsuit.* Available at: https://variety.com/2022/digital/news/facebook-90-million-privacy-lawsuit-settlement-1235182172/.

Stempel, Jonathan. *Zoom reaches $85 mln over user privacy, 'Zoombombing'.* Available at: https://www.reuters.com/technology/zoom-reaches-85-mln-settlement-lawsuit-over-user-privacy-zoombombing-2021-08-01/.

The Federal Trade Commission Act. 15 USC 41-58, as amended.

The YouTube Team. *Updates to YouTube's Terms of Service.* Available at: https://blog.youtube/news-and-events/updates-to-youtubes-terms-of-service/.

Turkewitz, Neil. *Fairness to Whom? Reimagining a New Paradigm for Considering Fair Use*. Available at: https://medium.com/@nturkewitz_56674/fairness-to-whom-reimagining-a-new-paradigm-for-considering-fair-use-c871797ceb60.

Uberti, David. *Fearing More Cyberattacks, Congress Requires Key Businesses to Report Digital Breaches.* Available at: https://www.wsj.com/articles/fears-of-cybersecurity-attacks-may-increase-disclosure-requirements-for-businesses-11647444384?mod=hp_lista_pos3.

Ziegler, Bart. *Should Amazon, Microsoft, Google and Other Cloud Companies Face More Government Oversight?* Available at: https://www.wsj.com/articles/should-amazon-google-microsoft-cloud-companies-face-more-government-oversight-11646430816.

8. Legal Case References

In re Clearview AI, Inc., Consumer Privacy Litigation, Case No. 21-cv-0135, Memorandum Opinion and Order (N.D. Ill. Feb. 14, 2022).

Outlogic, LLC v. NybSys, Inc., Case No. 21-cv-09592-VKD (N.D. Cal. 2021), First Amended Complaint.

The Andy Warhol Foundation for the Visual Arts, Inc. v. Lynn Goldsmith, Lynn Goldsmith, Ltd., 11 F.4th 26 (2d Cir. 2021).

.