AN INADEQUATE BAND-AID: EXISTING PRIVACY LAW HAS UNCERTAIN APPLICATION TO WEB-SCRAPED PERSONAL INFORMATION USED TO TRAIN AI

Jody L. Eckman, M.S. NCR

AN INADEQUATE BAND-AID: EXISTING PRIVACY LAW HAS UNCERTAIN APPLICATION TO WEB-SCRAPED PERSONAL INFORMATION USED TO TRAIN AI.

Jody L. Eckman, M.S. NCR*

I. INTRODUCTION

To download this very piece of writing, to turn on a "study" music playlist, to shop online for foam earplugs, or even map out a route to a local coffee shop—any one of these everyday tasks will produce a myriad of data. The more an electronic device or online service is used, the more data is generated. Multiply that individual data by 335 million people.[1] Then, imagine a patchwork band-aid of laws being all that exists to protect those consumers who embrace this new age of data. With inevitable wear and tear brought on by the exponential growth of data and technology, it is only a matter of time before this hypothetical legal band-aid's adhesive deteriorates.

Six years ago, it was estimated that 2.5 quintillion bytes[2] of data were being created every day by 3 billion internet users.[3] As of 2021, the amount of data available only swelled: an added 21.7%[4] of the world's population has moved online. What is to be made of this data stockpile? Depending on how often "study" music is played, Spotify will use that listening data to include the songs in the user's annual "Spotify Wrapped."[5] And if enough time is spent in one particular location, Apple or Google Maps might use the data to suggest a route back "home" after that afternoon coffee has been picked up.[6] With the earplug-purchase data from Target, coupons for highlighters and notebooks might arrive, or perhaps for some band t-

* J.D. Candidate at Marquette University Law School, class of 2024; B.A., Creighton University; M.S. NCR, Creighton University. I would like to thank Professor Bruce Boyden for his patience, feedback, and valuable insights during the research and writing process for this paper, along with countless family and friends for their unwavering support.

1. *U.S. and World Population Clock*, U.S. CENSUS BUREAU, https://www.census.gov/popclock/ (last visited Aug 9, 2023).

2. Bernard Marr, *How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read*, FORBES (May 21, 2018,12:42 AM), https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/?sh=2d91ed8160ba.

3. Brahima Sanou, *The World in 2014: ICT Facts and Figures*, INT'L TELECOMM. UNION (Apr. 2014), https://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2014-e.pdf [https://perma.cc/CZ7E-TDZG].

4. Simon Kemp, *A Decade in Digital*, DATAREPORTAL (Nov. 29, 2021), https://datareportal.com/reports/a-decade-in-digital [https://perma.cc/X8WR-BF5T].

5. A feature of the online music-streaming platform, Spotify, "Spotify Wrapped" is a marketing campaign that enables users of the platform to view a user-friendly summary of data about their activity over the course of one year.

6. Brooke Nelson, *Significant Locations: How your iPhone Knows Where You've Been*, READERS DIGEST (May 10, 2023), https://www.rd.com/article/iphone-feature-tracking-location/ [https://perma.cc/R9WZ-UNRV].

shirts and drumsticks.[7] Over the years, businesses and data brokers have learned to make sense of raw data, harnessing it for use in sales and recommendations.[8] Now, artificial intelligence ("AI") providers are taking it a step further, creating services and products out of the data they have stockpiled.[9]

AI, which is both powered by data and creates data itself, has captivated everyday consumers and countless industries with the novel goods that have already been propagated. Just ask Siri a question or take a self-driving car for a spin, optimize prices and product recommendations, or implement a no-human-needed phone tree in a call center. While the aforementioned may now be commonplace AI technology, they are just the start of what AI providers can do when there is an abundance of data available to draw from. That, and rake in colossal profits. Less than a year after ChatGPT's launch, the product is already estimated to have reached $100 million[10] in annualized revenue. Further, forecasts anticipate AI to generate massive profits, some predicting up to $7.9 trillion per year.[11]

With such exponential growth promised by AI, monetary and otherwise, what then, will ensure its development transpires conscientiously for U.S. consumers? Existing consumer privacy laws might seem to be an obvious answer; however, as will be discussed below, how current regulations read—somewhat similar to whack-a-mole attempting to control specific methods by which data is collected—they serve more as a "band-aid." Controversies as to what portion of this online data treasure-trove can then be used, and what purposes it can be used for, are thus largely undetermined. Despite the transformative potential presented by AI providers, allowing their technology to develop in an unregulated manner would be irresponsible and discourteous to consumers; therefore, lawmakers must enact protective measures that grant both AI providers and consumers the ability to prosper in this new age of data.

This paper seeks to answer what liability AI providers, whose training datasets were compiled by either scraping data from websites or collecting the data directly from consumers, may face under existing privacy laws—specifically, those state laws enacted by Virginia, California, Washington, and Illinois, as well as the "Wiretap Act" at the federal level. In addition, this paper will propose a standard on which future legislation can be based to more adequately regulate AI to protect consumers' personal information.

---

7. *Interest-based Advertising*, TARGET, https://www.target.com/c/interest-based-advertising/-/N-ztavm (last visited Jan. 28, 2024) [https://perma.cc/WUV9-BWWK].

8. Nik Froehlich, *The Truth In User Privacy and Targeted Ads*, FORBES (Feb. 24, 2022), https://www.forbes.com/sites/forbestechcouncil/2022/02/24/the-truth-in-user-privacy-and-targeted-ads/?sh=3e7926e3355e.

9. *See What is AI as a Service*, RUNAI, https://www.run.ai/guides/machine-learning-in-the-cloud/ai-as-a-service (last visited Jan. 28, 2024) [https://perma.cc/WUV9-BWWK].

10. Matt Bornstein, Guido Appenzeller & Martin Casado, *Who Owns the Generative AI Platform?*, ANDREESSEN HOROWITZ (Jan. 19, 2023), https://a16z.com/2023/01/19/who-owns-the-generative-ai-platform/ [https://perma.cc/3NGN-RZF9].

11. Francesco Guerrera, *AI's Deflationary Winds Will Blow Away Profits*, REUTERS (Jun. 28, 2023, 4:47 AM), https://www.reuters.com/breakingviews/ais-deflationary-winds-will-blow-away-profits-2023-06-27/.

The following discussion will proceed in three parts. Part I will lay out the background of AI and how it has developed: what the technology is, how it works, and its current capabilities. Part II will analyze how existing laws apply to AI providers, concluding that personal data collected by web-scraping may proceed with simple caution to regulations, but where the same personal data is collected directly from users, then more caution is advised, especially in the realm of obtaining adequate consent. Part III offers a proposal for how legislators might regulate AI providers to protect consumers. More specifically, lawmakers would be wise to institute controls, such as incorporating a set of principles similar to those found in GDPR along with duties of care and loyalty, for how AI providers handle personal data post-collection.

II. BACKGROUND

"Data" can take on many meanings—for some it might limit how much internet a phone can use, others see it as the new "oil" for a digital world take-over, and yet there are those who may not know it even exists.[12] Without further insight, the accepted definition might be deceptively boring ("facts and statistics collected together for reference or analysis")[13] yet in reality, what can be done on account of these nine words has sparked a new era in technology. Data is the fuel that is propelling AI. The more data AI providers can collect, the more powerful their products and services can become, drawing more users who subsequently generate more data.[14] This data life cycle may seem circular, but the spiral it creates is one of innovation. With increased access to data, many industries can be propelled to better themselves, and in turn, society too.

If data is the fuel, then AI is the engine that makes possible the commute to novel modernization. AI encompasses an umbrella of technology. At its core, though, the term generally refers to a computer program that can generate responses that were not provided in advance.[15] AI in this sense has been around for some time now. Amazon's Alexa voice assistance, Apple's facial recognition as a method to unlock an iPhone or verify a purchase, and Microsoft and Google's ability to separate incoming emails between "focus," "other," and "junk" inboxes, are just a few examples of machines functioning with foresight in their environment. Once glamorous, state-of-the-art AI forms, these technologies have since been widely adopted and are now used without a second thought. But for any of these aforementioned AI technologies, there are equally new and exciting AI tools, like

12. Kiran Bhagesphur, *Data Is The New Oil -- And That's a Good Thing*, FORBES (Nov. 15, 2019), https://www.forbes.com/sites/forbestechcouncil/2019/11/15/data-is-the-new-oil-and-thats-a-good-thing/?sh=38afb94b7304; Michelle Faverio, *Share of Those 65 and Older Who are Tech Users has Grown in the Past Decade*, PEW RSCH. CTR. (Jan. 13, 2022), https://www.pewresearch.org/short-reads/2022/01/13/share-of-those-65-and-older-who-are-tech-users-has-grown-in-the-past-decade/.

13. *Data*, OXFORD ENGLISH DICTIONARY, https://www.oed.com/dictionary/data_n?tab=factsheet#219838686 (last visited Aug 12, 2023).

14. *The World's Most Valuable Resource is No Longer Oil, but Data*, THE ECONOMIST (May 6, 2017), https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data [https://perma.cc/XU7A-NBJ5].

15. *See What is Artificial Intelligence (AI)?*, GOOGLECLOUD, https://cloud.google.com/learn/what-is-artificial-intelligence (last visited Jan. 28, 2024).

chatbots or virtual assistants, which incrementally build on previous AI. This progression in AI is thanks to the increased stash of data—that stash originating from consumers and society at large.[16]

There are few bounds to what can be done with data when it is combined with AI, but this paper limits its discussion to AI that puts data to work through subcategories of machine learning, deep learning, and large language models ("LLMs"). To start with AI itself: this umbrella category generally involves applying advanced analysis to logic-based techniques.[17] How exactly a machine "learns" or "thinks" depends on the technique at hand. Outputs produced by the machine learning subcategory involve taking in mathematical models and finding patterns to create algorithms or statistical formulas that can convert information into a single, predictive result.[18] A simple illustration of machine learning is customized user interfaces, such as recommended shows and movies that Netflix offers to a user after they have watched or interacted with the platform.[19] Deep learning is comparable to machine learning, but instead, this subcategory relies on multiple layers of information and the transformation of content at every level.[20] This type of AI is best where there exists a need to arrive at high-accuracy solutions despite an increase in the complexity of the problems; for instance, self-driving cars use deep learning as they must be able to safely navigate busy roadways, different forms of weather and road conditions, as well as varying groups of pedestrians.[21] LLMs on the other hand use a text-oriented framework.[22] Popularized by ChatGPT in 2022, this type of AI is trained on and produces outputs that are based on data collections comprised of billions of words.

For any AI subcategory to be of use to consumers, the AI product first needs to be "trained." Analogous to a human brain which might draw on reading materials or first-hand experiences, AI must amass information from some source to similarly produce its output.[23] AI cannot reflect on experiences or learn in the same way humans can, but it can learn from what humans create: data.[24] What an algebra textbook can teach a student is what a training data set can teach AI.[25] AI training data sets contain an assortment of information that is labeled in various ways and

16. Samir Sampat, *Where Do Generative AI Models Source Their Data & Information?*, SMITH.AI (Sept. 20, 2023), https://smith.ai/blog/where-do-generative-ai-models-source-their-data-information#:~:text=Web%20scraping%20and%20crawling,find%20the%20information%20they%20need [https://perma.cc/C9PT-VF58].

17. *What is Artificial Intelligence?*, GARTNER (2023), https://www.gartner.com/en/topics/artificial-intelligence.

18. *Id*.

19. *How Did Netflix Use ML to Become the World's Streaming Leader?*, (Feb. 8, 2022), https://dev.to/mage_ai/how-did-netflix-use-ml-to-become-the-worlds-streaming-leader-b3e#:~:text=Customizing%20user%20interface,-Once%20Netflix%20gets&text=To%20achieve%20success%20in%20targeting,uses%20machine%20learning%20(ML) [https://perma.cc/5RMP-DMYJ].

20. *What is Artificial Intelligence?*, *supra* note 17.

21. *Id*.

22. *Id*.

23. *See* Amal Joby, *What Is Training Data? How It's Used in Machine Learning*, LEARNG2 (July 30, 2021), https://learn.g2.com/training-data [https://perma.cc/38SN-SCUH].

24. *See id*.

25. *See id*.

arranged into categories.[26] Therefore, to give the AI training data is to "teach" it new information. Once the AI has consumed the training data, that data set may be stored, but is not specifically called on or pulled from as the AI produces later outputs.[27] This means for AI providers to build their own "brain" which the product or service will operate off of, they must (i) collect or source enormous quantities of data on an infinite range of topics, (ii) feed their AI the information in the form of a training data set so that it may extract knowledge, and (iii) apply its particular technique to the digested information as outputs are later produced.[28]

Data sets used for training will typically originate from a collection of sources, but those of interest to this discussion include information scraped from the open internet as well as data created by consumers as they use the AI product or service themselves, referred to as "user inputs."[29] To "scrape" data involves the use of bots to find and index information found on public-facing webpages.[30] For context, online sites that have been scraped for training data include Wikipedia, GitHub, news sites such as the New York Times and the Washington Post, image datasets such as Flickr and Getty, advertisements, personal blogs, government sites such as voter registration databases and real estate records, patent indexes, and social media sites that do not require an account or password to have access, such as Reddit.[31] Applied to a well-known form of AI previously mentioned, Amazon's Alexa received training from data scraped across hundreds of billions of these types of webpages. Doing so enabled "Alexa" to understand what language a consumer spoke, as well as to respond with helpful, related information no matter the wide range of questions that could be posed by consumers.

Collection of user inputs can be more straightforward than scraping. On a basic level, this data collection practice simply involves tracking what a user does—keystrokes, clicks, time spent on any given page or information, etc.—and cataloging that information. Microsoft and Google's email sorting AIs represent a form of collection directly from a user. Data is collected and cataloged as a user opens, deletes, or performs other actions on the messages in their inbox. Over time the AI learns to preemptively sort future messages to different inboxes for the user based on their past behaviors and interactions with similar messages.

The forms of AI discussed thus far largely fall on the elementary end of the AI risk spectrum; nevertheless, AI developers, businesses who use AI, and lawmakers cannot afford to overlook the potential harm that could surface from any point on the

---

26. Kate Crawford and Trevor Paglen, *Excavating AI: The Politics of Training Sets for Machine Learning*, (Sept. 19, 2019), https://excavating.ai [https://perma.cc/Y8YV-GG4G].

27. *See* Amal Joby, *What Is Training Data? How It's Used in Machine Learning*, LEARNG2 (July 30, 2021), https://learn.g2.com/training-data [https://perma.cc/R9CM-RRFE].

28. *See id*.

29. Samir Sampat, *Where Do Generative AI Models Source Their Data & Information?*, SMITH.AI (Sept. 20, 2023), https://smith.ai/blog/where-do-generative-ai-models-source-their-data-information#:~:text=Web%20scraping%20and%20crawling,find%20the%20information%20they%20need [https://perma.cc/5BFV-QR6A].

30. *See id*.

31. Kevin Schaul, Szu Yu Chen, & Nitasha Tiku, *Inside the Secret List of Websites that Make AI Like ChatGPT Sound Smart*, THE WASH. POST (Apr. 19, 2023, 6:00 AM), https://www.washingtonpost.com/technology/interactive/2023/ai-chatbot-learning/.

continuum that is AI technology. The risk most related to consumer's personal information is security—nearly every AI chatbot that has been released for use, at the time of this writing, has disclosed confidential information.[32] This type of leak is a serious concern from a legal perspective, but especially so on a practical level for individuals whose biometric information is involved given, for example, one's inability to alter their eye should data about their iris be exposed. Another risk both AI providers and users should be cognizant of is overreliance on the technology: where outputs are not checked against any other information or standards, harm can occur and trickle down to any number of people. For instance, an attorney in New York who used AI to conduct legal research found himself being reprimanded for doing so as the AI "hallucinated," citing to made-up legal precedent in its outputs.[33] Further, where AI-based decisions are being implemented in businesses, the information consumers provide could come back to haunt them if AI evolved to incorporate biases in its outputs that discriminate based on race or sexual orientation, among other means.[34]

Despite the liability AI exposes businesses and consumers to, it would be a mistake to do away with or severely limit AI's potential. With appropriate safeguards in place, AI can promote freedom, equality, and transparency. In healthcare settings, AI-based applications could improve health outcomes and the quality of life for millions of patients.[35] In education, AI can enhance lessons by personalizing instruction based on a student's learning style.[36] For financial institutions, AI could also be fashioned to readily detect and prevent cases of fraud.[37] In the legal field, AI has the potential to be especially conducive to research, document drafting, litigation analysis, and more, allowing parties to increase efficiency and derive more value from interactions.[38]

Advantageous or injurious, with few barriers and seemingly limitless potential, AI will continue to "learn" from the ever-expanding data reservoir and create new information, tools, and solutions. Before AI, general technology was already developing at exponential rates which legislation could not keep up with.[39] Given an

---

32. Sayash Kapoor & Arvind Narayanan, *A Misleading Open Letter About SCI-FI AI Dangers Ignores the Real Risks*, (Mar. 29, 2023), https://www.aisnakeoil.com/p/a-misleading-open-letter-about-sci [https://perma.cc/6A52-W4ZZ].

33. Sara Merken, *New York lawyers sanctioned for using fake ChatGPT cases in legal brief*, REUTERS (June 26, 2023), https://www.reuters.com/legal/new-york-lawyers-sanctioned-using-fake-chatgpt-cases-legal-brief-2023-06-22/.

34. Peter Stone, et.al., *Artificial Intelligence and Life in 2030*, One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel, STAN. UNIV. (Sept. 6, 2016), https://ai100.stanford.edu/sites/g/files/sbiybj18871/files/media/file/ai100report10032016fnl_singles.pdf [https://perma.cc/XZM7-PGX6].

35. *Id*.

36. *Id*.

37. *Id*.

38. *The Power of Artificial Intelligence in Legal Research*, LEXISNEXIS (May 16, 2023), https://www.lexisnexis.com/community/insights/legal/b/thought-leadership/posts/the-power-of-artificial-intelligence-in-legal-research [https://perma.cc/Z2MB-WY2A].

39. Manav Tanneeru, *Can the Law Keep up with Technology?*, CNN (Nov. 17, 2009), https://perma.cc/PAN9-HU6Z Legal experts state how difficult it is for the law to keep up with technology); *see also* Marci Harris, *Here's What Happens When Tech Outpaces Government*, APOLITICAL (Sept. 12, 2019),

implicit ability to create new data and solve problems on its own, AI, however, is unlike any other technology policymakers have seen before. This rate of growth should serve as an incandescent red flag. What the technology offers—beneficial and detrimental—compels serious consideration for policymakers.[40]

III. ANALYSIS

A. AI Trained on Data Scraped from Public Sources

AI necessitates access to an enormous pool of data for training. One source that supplies this data basin is publicly available websites. While any number of legal issues might arise from the practice of scraping websites, the concern at issue here is the privacy of personal information garnished as data is scraped. AI training data sets that are compiled by scraping may include consumers' personal information and therefore may be subject to various data privacy laws. The following section will analyze what AI providers might expect as they scrape data in four different states that are considered to have influential consumer privacy laws: Virginia, California, Washington, and Illinois.

i. The Virginia Consumer Data Protection Act

Virginia's Consumer Data Protection Act (the "VA CDPA") sets out to protect consumers' personal information. However, the legislation was written with such broad and numerous exemptions that AI providers will likely be able to avoid the Act altogether. While individuals in the state are equipped with a series of rights regarding their personal information, those rights only go so far because the scope of the Act is limited to companies' use of "personal data."[41] Virginia defines "personal data" as "any information that is linked or reasonably linkable to an identified or identifiable natural person," but goes on to exclude "publicly available information."[42] This single exception is what forms a wide avenue for AI providers to scrape large amounts of data from the internet. More specifically, the Act's text construes "publicly available information" to include not only the information that is "lawfully made available through federal, state, or local government records," but also any information which a business might have a "reasonable basis" to believe the consumer made publicly available via widely distributed media.[43] That is to say, the only exception to this "publicly available information" exception is the scenario

https://perma.cc/4CYX-FFSH (illustrating that the pace of development is an issue given how slow policy change occurs).

40. *See e.g.*, *US: Congress must regulate artificial intelligence to protect rights*, HUMAN RIGHTS WATCH (Oct. 17, 2023), https://www.hrw.org/news/2023/10/17/us-congress-must-regulate-artificial-intelligence-protect-rights?gad_source=1&gclid=CjwKCAiAk9itBhASEiwA1my_61CA7mBcjkvLQTUzk4eoWoOrg7vEutl-wwGwFPMXLa0DsQdVLJedGxoCKUcQAvD_BwE [https://perma.cc/Y6V5-LDKZ].

41. Va. Code Ann. §§ 59.1-576.

42. *Id.* §§ 59.1-575.

43. *Id*.

where a consumer takes steps to restrict the audience with whom their information is shared.[44]

Candidly, then, AI providers who scrape data only from public websites are unlikely to find themselves subject to the Act by way of the "publicly available information" exception, and consumers are thus left with a band-aid's worth of protection. Particularly relevant is that Virginia courts have yet to hear any case that would require them to demarcate websites being private versus publicly available. This is likely on account of the internet being generally understood as a public space where any user can search and find information when there is no prerequisite for a password or login. By the same token, neither are there any known cases that parse what "reasonable basis" a business must have to support its belief that a consumer's post was made public through widely distributed media.

Two lines generally can be crossed in which information on a website would then move from "publicly available" to "restricted" or "limited access." The first threshold materializes when an individual must create an account to use the website. For example, searching for the name "Jane Doe" on Google might result in links to social media pages with usernames or biographies that include "Jane Doe." An Instagram profile, for instance, might appear for someone named "Jane Doe" and allow the searcher to view two or three posts from that profile. However, to access more than this first impression the searcher likely needs to create their own Instagram account. Further, even by creating an account, the "Jane Doe" profile may still be inaccessible if its account owner has implemented the other general method to limit access: activating the "private account" setting(s). While options to restrict access to posts may differ per website, it is typical for there to exist some manner in which account owners can restrict what the public eye would otherwise be able to freely view. That said, imagine that one did create an Instagram account, but the content on Jane's profile still could not be viewed. This would likely be on account of Jane's choice to turn on her privacy settings—now only those users she approves of can interact with the information she chooses to share on her profile.

It is unlikely that AI providers scrape their training data from locations that require an account or are protected by user-implemented privacy settings. Thus, even if a Virginia court did define the outer limits of what is reasonably understood to be available to the general public, it would not likely include the sort of data included in AI training data sets. Moreover, web scrapers and AI providers are not new to Virginia; these entities were conducting business in the state at the time VA CDPA was written and passed. Therefore, if Virginia lawmakers intended for AI providers' current scraping practices to be subject to this legislation, they could have formulated the law to achieve that end.

ii. The California Consumer Privacy Act

Compared to Virginia, the California Consumer Privacy Act ("CCPA") is thought to be one of the most comprehensive pieces of data privacy legislation for consumers; however, to the extent that AI providers only scrape their training data

44. *Id*.

from publicly available websites, they can anticipate being free of liability under the CCPA as well. To find themselves subject to the CCPA, a California business must, among other requirements and thresholds, collect consumers' "personal information," which encompasses any information that, "identifies, relates to, describes, is reasonably capable of being associated with, or could reasonably be linked, directly or indirectly, with a particular consumer or household."[45] Akin to Virginia, California also exempts "publicly available" information from that which would otherwise fall under "personal information."[46] The bounds of "publicly available" again include certain government records, information a business reasonably believes the consumer made public, as well as that which is made available by someone the consumer disclosed the information to without restriction.[47] Moreover, recent litigation in California found that information on social media site Meta that could be viewed without first logging into the site was assumed to be publicly available and thus eligible to be scraped without violating any law or terms of service.[48] This recent decision, together with the text found in the CCPA, therefore sends a message to AI providers that they may scrape personal information from websites without liability attaching, so long as those websites and the information they scrape from them would be considered "publicly available."[49]

Whereas both the CCPA and VA CDPA provide avenues for certain AI providers to fall outside each Act's scope, AI providers should nonetheless take caution of the reasonableness standard incorporated in each Act's respective "publicly available" exception. Such a standard gives consumers a foot in the door to litigate what constitutes a reasonable belief that a website is available to the public. Otherwise said, if Virginia's "band-aid" is generic, the California "band-aid" might be a better-known name-brand. With this in mind, AI providers would be well-advised to take proactive measures such as documenting the availability of those websites they scrape at the time scraping occurs. Despite the minimal effort likely needed to prove a reasonable belief in the public nature of any personal information collected, precautionary measures further ensure AI providers engaged in scraping will have one less regulation to worry about.

iii. The Washington My Health My Data Act

With a special focus on health data privacy rights, Washington state has also passed its privacy law, and, despite the extensive coverage provided to consumers in the state, the act is similar to that of California and Virginia: AI providers, with caution, are once again likely not subject to the legislation. Only those entities in Washington that deal with "consumer health data" are subject to the My Health My Data Act ("WA MHMD"), which, congruent with the CCPA and VA CDPA, similarly incorporates an exclusion for "publicly available information."[50] Moreover,

45. California Consumer Privacy Act, Cal. Civ. Code § 1798.140 (v)(1).

46. *Id*. § 2(v)(2).

47. *Id*.

48. Meta Platforms, Inc. v. Bright Data Ltd., No. 23-CV-00077-EMC, 2024 WL 251406, at *17 (N.D. Cal. Jan. 23, 2024).

49. *See* Cal. Civ. Code § 1798.140 2(v)(2).

50. Washington My Health My Data Act, Wash. Sess. Laws Sec. 3(18)(b).

the definition of "publicly available information" is synonymous with that of VA CDPA and the CCPA as it includes data located in government records or widely distributed media, as well as that where there is a reasonable basis to believe the consumer made the data available to the general public.[51]

Where an individual consumer has chosen to share their information online without any privacy settings that would otherwise limit who can engage with that content, it can be assumed there exists a reasonable basis for a business to believe that the individual consumer lawfully made that specific information available to the general public. This means that once more, no matter the later handling practices or security measures taken, AI providers who, for training purposes, collect data by way of scraping publicly available websites, will likely be discharged from the legislation's demands.

iv. The Illinois Biometric Information Privacy Act

Illinois' legislation highlights a faction of consumer privacy law different from that of Virginia, California, and Washington—here, lawmakers emphasize consumer privacy protection based on biometric identifiers and biometric information. Despite this divergent focus, the Biometric Information Privacy Act ("BIPA") is nevertheless like that of the previously mentioned states: AI providers in Illinois who scrape data for training purposes but do not scan biometric identifiers or biometric information in the process, can, with caution, proceed yet to collect the data.

AI providers whose outputs include AI-generated images will have to be particularly careful, however. The main issue these types of AI providers will need to consider is whether their image-generation tools construct a "biometric identifier" or constitute "biometric information" under BIPA.[52] Illinois' BIPA defines "biometric information" as any information "based on an individual's biometric identifier used to identify an individual."[53] This means that information drawn from a scan of a person's facial geometry is "biometric information" under BIPA. The Act goes on, however, to omit "photographs" from biometric identifiers, which indicates that information extracted from a photograph is excluded from the definition of "biometric information" unless that information squarely qualifies as a "biometric identifier," e.g., the information measures an individual's facial geometry.[54] In total, so long as AI providers are not extracting facial geometry or other biometric identifiers from photographs, they will not be subject to BIPA.

It is clear that photographs alone are neither biometric identifiers nor part of biometric information;[55] however, where a photograph is scanned for biometric identifiers, such as facial geometry, the data that is extracted is considered biometric information and will subject the scanner to BIPA.[56] In *Monroy*, the plaintiff's friend uploaded a photograph of the plaintiff to Shutterfly, a digital photograph storage website. Shutterfly then, without the plaintiff's consent, scanned the photo,

51. *Id*. § 3(22).
52. Illinois Biometric Information Privacy Act, 740 ILCS §§ 14/5(c), 10 (2023).
53. *Id*. § 10.
54. *Id*.
55. Monroy v. Shutterfly, Inc., No. 16 C 10984, 2017 WL 4099846, at *3 (N.D. Ill. Sept. 15, 2017).
56. *Id*. at *5.

evidenced by Shutterfly's prompt to the uploader to tag the face shown in the photograph.[57] The court found that such extraction of face geometry did indeed constitute a biometric identifier and therefore found Shutterfly to be subject to BIPA.[58]

While AI providers are similar to the defendant in *Monroy* due to their use of photographs without the pictured individual's consent, what will determine whether BIPA applies is what the AI service provider does next. The defendant in *Monroy* extracted facial geometry from the user-uploaded photographs to tag other images featuring the same person; AI providers, on the other hand, are only attempting to train their algorithm, which may or may not require the use of a particular individual's face. This is of relevance as some AI-generated images have produced images that feature recognizable celebrities in novel situations. If an AI provider were to scan the facial geometry extracted from the photographs, such as to create additional images of the same person or similar-looking people, then they would find themselves in the shoes of the defendant in *Monroy*: subject to BIPA. But if AI providers scrape photographs from websites to teach their algorithms only to create entirely new faces without scanning biometric identifiers, then BIPA will likely not apply to those AI providers.

In sum, where the information defined in the privacy legislation is not collected by AI providers, it is unlikely any culpability will attach. Thus, Illinois' BIPA joins the growing list of state privacy laws considered above that may not apply to AI providers' products and services that are trained on scraped data.

B. AI Trained on User Data

Vast as the internet is, AI providers need not comb websites to find valuable data; training data can be sourced internally, too. A curious mind may have noticed over the past several years that an increasing number of products and services now prompt users with terms or privacy notices that mention the consumers' inputs may be "documented" and "later used" by the business. Upon solicitation or use of an AI product or service, AI users themselves can reduce the AI provider's need to source data externally given they provide ample data as they use the product or service—that data which may be more beneficial given its direct relation to the AI product or service. Amazon's privacy notice illustrates this phenomenon as it details how the company, "automatically collect[s] and store[s] certain types of information about your use of [our] services" in order to "improve [our] products and services."[59] Awareness of such user data collection yet no clear-cut details as to what is being done with that data, therefore, begs the questions: how is personal data being handled, and has the user indisputably consented to that handling?

To summarize the foregoing discussion, though AI providers are likely in the clear when it comes to their tactic of scraping public websites for training data, when the source of training data is the user themselves, then AI providers will almost certainly be subject to state consumer privacy laws. With no exemption lifeboat in

57. *Id.* at *1.
58. *Id*. at *5.
59. *Privacy Notice*, AMAZON (Aug. 11, 2023), https://aws.amazon.com/privacy/.

sight this time around, AI providers, thus, need to also prepare to answer to their new responsibilities found under these state consumer privacy laws. Further, as the collection of user data involves intercepting electronic communications, AI providers should thus be cautious of the federal Wiretap Act as well.

i. State Consumer Privacy Law Applied to User Data

Under the lens of data collected straight from users, a user, whether by direct disclosure or by interacting with a product or service, makes their personal information available in a presumably "private" setting. None of the earlier mentioned states— Virginia, California, Washington, or Illinois—detail an exception for sourcing data in this manner. The only mention of consumer-shared personal information being exempt is in the case where disclosure occurs in a *public* manner, which is opposite to the current inquiry. Thus, with no lifeboat of an exemption in sight, each state's consumer privacy law, as outlined below, will very likely require the attention of AI providers as they collect personal information through this internal user data source.

Virginia's CDPA considers "personal information" to include anything that could be "linked" to an identified, natural person.[60] Where a consumer provides data, say through setting up a profile where the user's date of birth, address, and email are all input, the AI provider will no doubt become subject to VA CDPA given the contents, alone or combined, could potentially be linked back to the consumer. Where the consumer, as they input this information, is not simultaneously sharing it in a public manner, the AI provider has no exemption to run to and will be subject to the entirety of VA CDPA.

This analysis follows even more readily for California whose "personal information" umbrella under the Act is that much wider, including information that not only identifies, like Virginia, a particular consumer or household, but also that which might, "relate[] to, describe[], is reasonably capable of being associated with, or could reasonably be linked, directly or indirectly [to] a particular consumer or household."[61] Where the information AI providers compile originates explicitly from their users, it is self-evident that data will then fall under the CCPA's "personal information" definition thus simultaneously subjecting the AI provider to this state law as well.

It should be no surprise, given the above analysis, that AI providers will find themselves subject to Washington's MHMD as well, despite the law's specific focus on consumer health data. Comparable to the text found in the CCPA, personal information under MHMD is similarly defined with a wide scope, encompassing both what "identifies" as well as that which is "reasonably capable of being associated or linked, directly or indirectly, with a particular consumer."[62] Thus, just as an AI provider is answerable to the CCPA when information is gathered straight from the AI provider's users, they are equally answerable to MHMD in Washington.

60. *Supra* note 22, § 59.1-575-585 (2023).
61. California Consumer Privacy Act, Cal. Civ. Code § 1798.140 (v)(1) (2023).
62. Washington My Health My Data, Sec. 3(18)(a).

Even where state law focuses on biometrics, such as that of Illinois' BIPA, no relief will be found for AI providers. BIPA makes clear that the method by which "biometric information" is captured will not dictate whether a business' is subject to the law.[63] Instead, it is the *type* of information collected that Illinois looks to regulate. More specifically, BIPA targets *any* collection source where biometric identifiers or biometric information is involved. Thus, the user-provided data method is irrelevant and will subject the AI provider to Illinois' BIPA so long as some piece of what the AI provider actually collected falls under "biometric identifiers" or "biometric information."

The commonality of user-provided data as a data source does nothing for AI providers in terms of their being subject to state privacy laws. Simply put, a consumer who has handed over the type of information the statute looks to regulate, whether "personal" or "biometric" information, will likely lead the AI provider to be subject to some piece of privacy legislation. Further, what comes along with being accountable to such privacy laws will generally be an array of consumer rights and thus potential litigation. More specifically, where a training data set contains personal information, a business needs to determine how they might (i) de-identify data before the algorithm is trained on it; or (ii) track the data that is already baked into the general knowledge of the algorithm so that it can be quickly pinpointed were a consumer to exercise their right to know; and (iii) additionally have the capability to locate data for the purpose of genuinely complying with consumers' requests to delete. The ability to accomplish the third task is imperative given the way data is initially baked into the AI product or service and algorithm as a whole. All said, AI providers ought to think far beyond initial collection from users when determining where they will source their data from.

ii. Federal Wiretap Act Applied to User Data

Beyond state law, AI providers who collect, use, and share user data will also want to consider their accountability under the federal Wiretap Act. This piece of legislation prohibits the intentional interception of any electronic communication where the person who has intercepted the communication is neither a party to the communication nor received consent from at least one person who is a party to the communication.[64] If AI providers have not, as compared to Amazon's earlier mentioned verbiage, changed the level of granularity with which they obtain consumers' consent to make use of the consumer's data generated while interacting with the business's products or services, then AI providers may risk prosecution or fines under the Wiretap Act for want of adequate consent.

AI providers who collect the interactive data users produce on their AI platform but have not first received clear consent from those users to collect that interactive data, may be exposed to liability under the Wiretap Act.[65] In *In re Facebook, Inc.*, where the defendant, Facebook, compiled data about the browsing history of its users without consent, the court held that, on the facts alone, the plaintiffs sufficiently

63. Illinois Biometric Information Privacy Act, 740 ILCS § 14/10 (2023).

64. 18 U.S.C. § 2511(2)(d) (2018).

65. In re Facebook, Inc. Internet Tracking Litig., 956 F.3d 589, 598 (9th Cir. 2020).

alleged an invasion of privacy under the Wiretap Act given Facebook's lack of obtaining clear consent from the users to use or share the information.[66]

Moreover, for actual consent to be found, the disclosure provided to users must have explicitly notified them of the practice at issue.[67] In *In re Google Inc.*, the defendant, who gathered information from the plaintiff as they sent email communications via defendant's services, argued that requisite consent was collected when the plaintiff agreed to the defendant's general terms of service and privacy policies.[68] More specifically, those terms stated, "advertisements may be targeted to . . . queries made through the Services or other information."[69] The court rejected this argument, however, and held that plaintiffs were not notified of the explicit information the defendant would intercept as part of their practice, nor the purpose of it, and thus, the defendant had not obtained sufficient consent.[70]

Recently, however, a California court held that a website operator was the known and intended recipient of communications sent using the AI chat feature, and, thus, was considered to be a party to the communication, alleviating the requirement for consent.[71] Given the many forms in which AI can be presented, it is likely that certain AI providers can similarly claim the party exemption under the Wiretap Act as well. Of note though are individual state laws that mimic the Wiretap Act to some degree and may require both parties' consent to the communication.[72] Thus, any AI provider who collects data from customer interactions would be wise to thoroughly outline how consent will be obtained.

Where AI providers are like that of the defendant in *In re Facebook*—intercepting the interactions of users and compiling that information—as well as that of the defendant in *In re Google*—detailing neither what specific data is collected as users interact with the product nor what specific purpose is served by such collection—special caution should be taken. Like that of the outcome in *In re Facebook*, such AI providers may reasonably be found to have violated the Wiretap Act, especially so given the presumption from *In re Google*, when they forego providing specific details when obtaining user consent.

AI providers who do choose to collect consumer user data may consider first reflecting on the changes Zoom Video Communications, Inc. ("Zoom") recently made to its privacy policy. The new verbiage gives the company broad rights to use "customer content" for "product and service development."[73] However, with an eye towards understanding that "consent" has been interpreted narrowly to require a specific description of what will be done with the user's communications, Zoom

66. *Id*. at 599.

67. In re Google Inc., No. 13-MD-02430-LHK, 2013 WL 5423918, at *13 (N.D. Cal. Sept. 26, 2013).

68. *Id*.

69. *Id*.

70. *Id*.

71. Pena v. GameStop, Inc., No. 22-CV-1635 JLS (MDD), 2023 WL 3170047, at *12 (S.D. Cal. Apr. 27, 2023).

72. *All Party (Two Party) Consent States*, RECORDING LAW (Sept. 17, 2022), https://recordinglaw.com/party-two-party-consent-states/ [https://perma.cc/B3PY-FXDD].

73. Smita Hashim, *How Zoom's Terms of Service and Practices Apply to AI Features*, ZOOM, (Aug. 7, 2023), https://blog.zoom.us/zooms-term-service-ai/.

opted to specifically mention that collected data may be used in furtherance of machine learning and AI tools.[74] Additionally, Zoom took steps to clearly notify its users when this type of data would be collected as well as how the feature that enables collection could be turned off by the customer.

While an update in language to service terms and privacy policies could overcome possible disclosure and consent issues, what is most troublesome for AI providers is the unpredictable nature of outputs. Not only does the Wiretap Act prohibit intentional interception of electronic communications, but it also precludes intentional disclosure. Thus, if personal information from the training dataset were somehow incorporated into the algorithm in an identifiable way and then subsequently disclosed via an output to a third party, AI providers would again be liable under the Wiretap Act.[75] Therefore, AI providers ought to give special attention to implementing proactive measures that will ensure their products and services will not contain any individual's personal information in the outputs they produce.

To sum up, legislation that regulates on the basis of *where* personal data is collected *from* will not suffice to protect consumers in the age of AI. AI providers who compile their training datasets by scraping publicly available websites are unlikely to see restrictions to their practices under existing privacy laws due to the nearly universal "publicly available" exception. Neither will AI providers be prevented from gathering personal information directly from consumers so long as they incorporate methods that respect consumers' rights such as access and deletion while also obtaining actual, explicit consent. Altogether, AI providers at present can presumably, after considering their organization's specific policies and procedures, continue to scrape without liability attaching. So too may they also directly collect data from users, as long as they receive adequate consent given the shape existing privacy laws have taken.

IV. PROPOSAL

Based on the above analysis, AI providers can be said to, with caution, freely scrape the internet for publicly available information. Despite the implication this may have for personal information that is swept up, this paper does not propose there be any limit on the act of scraping itself. Rather, focus should be placed on handling of data once it has been scraped. Entities who scrape data should be obligated to process data and develop products in a conscientious and transparent manner.[76] AI providers must be motivated to create value for society.[77] Where there is regulation of AI providers, it ought to be on the basis of encouraging helpful innovation, generating and transferring knowledge, and fostering broad corporate and civic responsibility to address critical societal issues AI technologies inevitably raise.[78]

---

74. *Id*.

75. 18 U.S.C. § 2511(2)(d) (2018).

76. Timnit Gebru, et.al., *Statement from the Listed Authors of Stochastic Parrots on the "AI Pause" Letter*, DAIR (Mar. 31, 2023), https://www.dair-institute.org/blog/letter-statement-March2023/ [https://perma.cc/86UR-5AZZ].

77. Stone, *supra* note 34..

78. *Id*.

On July 21, 2023, AI providers in the U.S. took a pledge to instill principles of safety, security, and trust as they develop AI.[79] While a step in the right direction, the risks and potential innovations of AI nonetheless suggest that more than a pledge be taken. Therefore, this paper recommends lawmakers embrace the following measures: (i) establish principles, with reference to General Data Protection Regulation ("GDPR"), and duties which AI providers should regard as they process data, and; (ii) establish a regulatory body with the power to (a) outline policies, procedures, and risk assessments AI providers must partake in, as well as (b) address complaints and violations of the aforementioned powers.

A. Recommendation: Establish Principles and Duties to Govern AI Providers' Processing and Handling of Data.

Legislation that targets how data is collected is largely a waste of time. Not only are such rules likely to become feeble as technology quickly advances, but just as probable is it that companies with ample resources will find, or invest in sourcing, other methods of collection[80] that circumvent any stated requirements. Therefore, to protect consumer's personal information, practical AI provider regulation must focus on *how* data is handled and processed.

One recommended way to do this is by adopting a set of principles that AI providers must exemplify in their operational practices—those practices that the principles apply to must include, at a minimum, any originating reason for which consumer personal information will be collected in addition to security measures taken for data post-collection. The European Union's GDPR offers precedent with six principles under which businesses can lawfully process data: (a) consent; (b) legal necessity; (c) protection of a person's vital interests; (d) public interest; and (e) legitimate interests pursued by the controller, except where the data subject's interests or fundamental rights and freedoms would override those of the controller.[81] Further, GDPR outlines the following principles for the handling of sensitive information, all of which U.S. policymakers are urged to incorporate in reference to the handling of consumer's personal information: (a) lawfulness, fairness, and

79. *Fact sheet: Biden-Harris Administration Secures Voluntary Commitments From Leading Artificial Intelligence Companies to Manage the Risks Posed by AI*, THE WHITE HOUSE (Jul. 21, 2023), https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/ [https://perma.cc/G4AU-2T86].

80. While this writing focuses on scraping and user input methods of data collection, the method of purchasing data is also relevant to regulating how data is handled is. Data brokers have been shown to sell sensitive data without consumer awareness—i.e., mental health data, geolocation data, financial details. Justin Sherman, *Data Brokers and Sensitive Data on U.S. Individuals: Threats to American Civil Rights, National Security, and Democracy*, DUKE SANFORD CYBER POLICY PROGRAM, https://techpolicy.sanford.duke.edu/wp-content/uploads/sites/4/2021/08/Data-Brokers-and-Sensitive-Data-on-US-Individuals-Sherman-2021.pdf (last visited Jan. 28, 2024) [https://perma.cc/9ZL3-R4ME]; Joanne Kim, *Data Brokers and the Sale of Americans' Mental Health Data: The Exchange of Our Most Sensitive Data and What It Means for Personal Privacy*, DUKE SANFORD CYBER POLICY PROGRAM (Feb. 2023), https://techpolicy.sanford.duke.edu/wp-content/uploads/sites/4/2023/02/Kim-2023-Data-Brokers-and-the-Sale-of-Americans-Mental-Health-Data.pdf [https://perma.cc/K3HC-6EFL].

81. *See* Commission Regulation 2016/679 of Apr. 5, 2016, General Data Protection Regulation, 2016 O.J. (L 119) 2, art. 6,.

transparency; (b) purpose limitation; (c) data minimization; (d) accuracy; (e) storage limitation; and (f) integrity and confidentiality.[82]

Separately, it would be wise for policymakers to consider adopting a formal duty of care and duty of loyalty for AI providers as well.[83] The International Association of Privacy Professionals believes such a standard of care already exists when accounting for the culmination of requirements provided by the patchwork of U.S. state laws; therefore, developing explicit duties to reflect this reality would only expound that which is already "status-quo" while removing any gaps or room for leniency. More specifically, the duty of care would encompass current common practices such as data protection assessments, vendor contracts and oversight, de-identifying data, along with other technical, organizational, and physical safeguards. A duty of loyalty, respectively, would prohibit uses of consumers' personal data that otherwise conflict with the consumers' best interests. In practice, this would require several changes from AI providers. First, it would require discontinuing the present notice-and-consent framework and implementing instead an affirmative consent approach. Next, it would require providing consumers with the "usual arsenal" of protections comprised of the right to access, correct and delete personal data, opt out of targeted advertising, and not be discriminated against based on protected characteristics. Finally, and arguably most important, the duty of loyalty would involve explicit requirements for data minimization, purpose limitation, privacy by design, and sensitive personal data use. Were policymakers to implement principles like those of GDPR previously mentioned as well, this duty of loyalty would serve to reiterate and reinforce those fundamental standards as being of utmost importance.

U.S. policymakers need not adopt the entirety of GDPR nor form an identical mirror copy of the principles found in GDPR, nor do they need to implement a whole cluster of new duties. Pressing for lawmakers on this side of the pond, however, is a need for something more than the current "band-aid" of privacy laws. The European Union's GDPR and business law's duties of care and loyalty can minimally serve as models for framing a sound, U.S.-specific construction of protections.

B. Recommendation: Form AI Regulatory Body

Given the enormous increase in data produced by internet users combined with the progress AI providers have already achieved, some form of increased regulation is inevitable. Regardless of whether policymakers take note of the foregoing recommendations, for any governance that is passed to be effective, there must be an informed regulatory body that has the capacity to enforce the legislation in a meaningful way. On the whole, it is proposed that the group sit within the FTC, similar to that of the Bureaus of Consumer Protection, Competition, and Economics; the members who serve in this function showcase proficiency in technology, data, or

82. *See id*. at art. 5.

83. Sam Castic & Anokhy Desai, *Addressing the Duty of Care in State Privacy Laws*, IAPP (Aug. 15, 2023), https://iapp.org/news/a/addressing-the-duty-of-care-in-state-privacy-laws/?mkt_tok=MTM4LUVaTS0wNDIAAAGNmSMjXw23j8xFH4LNQjPrho4sJQ31J6dupe2EJ_9wXkIXRPB5_g0NXi1CAZlLS0ICOcRECV_3z_HMqJHJH8wn2_NE3e8B61d68F8guMsbxzwjFQ (explaining the duty of loyalty and a duty of care are valuable in a privacy context) [https://perma.cc/74V5-S4ET].

AI practices; and finally, the group itself be given the requisite powers to actually effectuate its prescribed role.

The group that is formed must be committed and adept to understand and analyze AI technologies, programmatic objectives, and overall societal values.[84] Should those appointed to regulate AI be without sufficient technical expertise, potentially promising applications of AI could be refused and therefore frustrate the progress of society as a whole.[85] On the contrary, were officials to greenlight a sensitive application without proper vetting, there could be, among many consequences, a collapse of necessary protection for sensitive consumer personal information.[86] Outright, were AI to be regulated by those individuals who have not yet formed a sufficient understanding of how AI systems interact with human behavior and societal values, society at large will be poorly positioned to build momentum and innovate with AI.

Further, it is recommended that this new commission be given the power to act in two different ways. First, they shall have the ability to institute policies, procedures, and risk assessments that AI providers must answer to. Such risk assessments should, at a minimum, require AI providers to investigate the likelihood that personal information will appear in output(s), and additionally, delve into and prevent users from manipulating the AI to leak other consumers' personal information. Moreover, the body shall serve as the dedicated liaison for complaints and violations. Those grievances they would hear include abuses of any principles, duties, and privacy legislation generally, as well as any instances of disregard for the above-mentioned policies, procedures, or risk assessments the agency has implemented.

In sum, even if policymakers were to pass the most balanced and well-written regulations, without a body that can understand the technology before them or wield any power to make those rules meaningful, the effort would be nothing more than a fool's errand, serving the likes of a generic-branded band-aid. For consumers and AI providers alike to thrive, both sound policies and bodies to oversee them are crucial.

V. CONCLUSION

ChatGPT may have initiated seismic waves in the last year, but in reality, the technology and infrastructure on which AI is built have already shaped society through decades-long evolution. Despite recent consumer privacy laws having been enacted, certain AI providers who scrape solely publicly available information may nonetheless cautiously maintain a reasonable belief that they can continue to operate as is without becoming subject to these new pieces of legislation. Whereas AI providers who collect personal information directly from consumers will more likely be subject to certain state and federal laws, their foremost concerns ought to be centered only on obtaining adequate consent and devising efficient and satisfactory methods to respond to consumers as they invoke their consumer privacy rights.

84. Stone, *supra* note 34.
85. *Id*. at 43.
86. *Id*.

AI possesses an inherent potential to transform life as we know it, leaving more than just a scrape in its path. Were lawmakers to continue to allow it to develop in a largely unregulated manner—based on particular existing privacy laws which are only superficial as applied to the practice of scraping their training data—it would be irresponsible and discourteous to consumers and their personal information; thus, lawmakers must create knowledgeable regulatory bodies and enact principles and duties that protect personal data, therefore allowing both AI technology and consumers to prosper in the new age of data. After all, a suture that ties together those elements of an open wound is better than 50 states of band-aids.