FIU Law Review

Volume 17
Number 2 *Privacy Discussion Forum – University of Stockholm (Stockholm, Sweden, 2022)*

Article 12

Spring 2023

Evolving Privacy Protections for Emerging Machine Learning Data Under Carpenter v. United States

Emily Nicolella
FIU College of Law, enico014@fiu.edu

Online ISSN: 2643-7759

Recommended Citation

Emily Nicolella, *Evolving Privacy Protections for Emerging Machine Learning Data Under Carpenter v. United States*, 17 FIU L. Rev. 453 (2023).

DOI: https://dx.doi.org/10.25148/lawrev.17.2.12

EVOLVING PRIVACY PROTECTIONS FOR EMERGING MACHINE LEARNING DATA UNDER *CARPENTER V. UNITED STATES*

Emily Nicolella[*]

ABSTRACT

The Fourth Amendment's third-party doctrine eliminates an individual's reasonable expectation of privacy in information they willingly turn over to third parties. Government scrutiny of this information is not considered a search under the Fourth Amendment and is therefore not given constitutional protections. In the 2018 case *Carpenter v. United States,* the Supreme Court created an exception to the third-party doctrine. In *Carpenter*, a case involving the warrantless use of cell site location information (CSLI) in a criminal investigation, the Court held that individuals do have a reasonable expectation of privacy regarding CSLI. According to Chief Justice Roberts, despite the necessary relinquishment of some information by all cell phone users, privacy is guaranteed "[i]n light of the deeply revealing nature of CSLI, its depth, breadth, and comprehensive reach, and the inescapable and automatic nature of its collection . . . ." The Court's rationale in distinguishing CSLI is also applicable to the personal data that is constantly being collected by tech companies through the use of machine learning algorithms. Companies like Facebook and Google use machine learning to specifically tailor each user's experience to their individual preferences. To do so, machine learning algorithms constantly collect, store, and analyze data about our interactions online to "learn" about our habits, ideologies, likes, dislikes, and affiliations. Given the *Carpenter* Court's understanding of the constitutional complexities of high-tech communications, this comment takes the next step to explore individuals' reasonable expectation of privacy in algorithmic learning data titrated to their personal preferences.

I. Introduction 454
II. Artificial Intelligence and Machine Learning 456
A. What Is an Algorithm? 457
B. Applications of Machine Learning by Modern Tech Companies 460

[*] J.D. Candidate 2023, FIU College of Law.

C. Uses of Machine Learning and Digital Data by Law Enforcement....463
III. The Fourth Amendment and Digital Data....465
A. The Trespass Framework....466
B. The Reasonable Expectation of Privacy Standard....467
C. The Fourth Amendment and Emerging Technology....469
D. *Carpenter v. United States* and Its Progeny....471
IV. Machine Learning Data Under *Carpenter v. United States*....472
A. The Intimate and Revealing Nature of the Data....473
B. Automatic and Continuous Data Collection....474
C. Prevalence in Modern Society....475
V. Conclusion....478

I. INTRODUCTION

In every facet of modern life, from the stock market to baseball to Facebook and beyond, complex computer algorithms are hard at work, tasked with distilling vast quantities of data, constructing digital models to analyze the past, and running simulations to optimize the future.[1] Algorithms are truly ubiquitous in modern society. From the most helpful and benign to the seemingly dystopian and sinister, the government and the private sector rely heavily on machine learning and algorithmic data analysis to perform a myriad of functions. Machine learning algorithms are behind the software that automatically routes emails to a junk folder or prescreens resumes and loan applications.[2] In the criminal justice system, computer algorithms are used to analyze crime data for use in policing and sentencing.[3] Using algorithms, sports teams have been able to detect patterns and predict outcomes by analyzing biometrics and game statistics for every player in every game, and this analysis has changed the way the games are played.[4] Algorithms do not just allow us to understand and solve modern problems but they change the way we approach and interact with the world.

[1] *See* CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY 6, 16–18 (Crown Books 2016); Michael L. Rich, *Machine Learning, Automated Suspicion Algorithms, and the Fourth Amendment,* 164 U. PA. L. REV. 871, 882–83 (2016); *see also* Kevin Slavin, *How Algorithms Shape Our World*, TEDGLOBAL (2011), https://www.ted.com/talks/kevin_slavin_how_algorithms_shape_our_world.

[2] O'NEIL, *supra* note 1, at 3; Rich, *supra* note 1, at 882.

[3] O'NEIL, *supra* note 1, at 23–27.

[4] *Id.* at 16–18; Nabeel Abdul Latheef, *The Number Games–How Machine Learning is Changing Sports*, MEDIUM (July 21, 2017), https://medium.com/@nabil_lathif/the-number-games-how-machine-learning-is-changing-sports-4f4673792c8e.

Some of the most pervasive and opaque machine learning algorithms are used by tech companies like Facebook and Google to track the online behavior of their users to tailor users' individual experiences, target advertisements, and maintain user engagement.[5] Through the use of machine learning algorithms, these companies collect and analyze immense quantities of data about its users' online habits across all associated platforms.[6] Google's algorithms, for example, collect, store, and analyze user data from Google's search engine, Gmail, Chrome, YouTube, and Maps.[7] Generally, government use of data collected by tech companies does not implicate the Fourth Amendment because of the third-party doctrine, which states that an individual has a reduced expectation of privacy in information they turn over to a third-party.[8] Under the traditional understanding of the third-party doctrine, in using these various applications, whether it be Instagram, YouTube, or Google Maps, the user "gives" their engagement data to the company, thereby relinquishing many Fourth Amendment protections.

In 2018, however, the Supreme Court acknowledged an exception to the otherwise rigid third-party doctrine.[9] Cell-site location information (CSLI), the Court reasoned, is distinct from other types of personal data due to its comprehensive nature, its immense probative value, and the constant nature of its collection.[10] Although the majority in *Carpenter* stated that their holding was narrow, Justices on the Court and legal scholars have contemplated the far-reaching applications of the *Carpenter* framework.[11] Additionally, lower courts have already started to rely on *Carpenter* to invalidate searches and seizures of personal data beyond that of CSLI, such as information posted to personal social media pages.[12]

A category of information that fits under *Carpenter*'s rationale is the user engagement data that is constantly being collected and analyzed by machine learning systems. This comment asks whether the Supreme Court's treatment of CSLI in *Carpenter* should extend to the vast quantities of personal data collected by the machine learning algorithms used by tech corporations. Part II of this comment explains the basics of traditional and

[5] Sarah Morrison, *Why You Should Care About Data Privacy Even if You Have "Nothing to Hide,"* VOX (Jan. 28, 2021, 1:10 PM), https://www.vox.com/recode/22250897/facebook-data-privacy-collection-algorithms-extremism.

[6] *Id.*

[7] *See id.*

[8] *See generally* United States v. Miller, 425 U.S. 435 (1976); Smith v. Maryland, 442 U.S. 735 (1979).

[9] Carpenter v. United States, 138 S. Ct. 2206, 2223 (2018).

[10] *Id.*

[11] *Id.* at 2233–35 (Kennedy, J., dissenting); Matthew Doktor, *Facial Recognition and the Fourth Amendment in the Wake of Carpenter v. United States*, 89 U. CIN. L. REV. 552, 567–68 (2021).

[12] Doktor, *supra* note 11, at 566–67.

machine learning algorithms and discusses how they are used by the government and modern tech companies. Part III explores the development of the Supreme Court's Fourth Amendment jurisprudence on digital information. It includes an analysis of the traditional trespass framework, the reasonable expectation of privacy standard, the third-party doctrine, and the Supreme Court's ruling in *Carpenter v. United States*. Furthermore, it surveys the application of *Carpenter* in the lower federal courts in cases involving digital data. Part IV discusses the similarities between CSLI and the personal data compiled by internet machine learning algorithms, and asks whether such data should be given increased constitutional protections given the Court's recent treatment of digital data under the Fourth Amendment.

II. ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

The vast arena of artificial intelligence includes any "automated, machine-based technologies with at least some capacity for self-governance that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments."[13] The process by which these systems become "intelligent" is called machine learning. Despite their prevalence in every facet of modern life, the concepts of artificial intelligence and machine learning have been muddied by their depiction in movies and television. In media, machine learning systems often serve as substitutes for roles that would otherwise be filled by humans. In the 2013 movie *Her*, a sensitive and introverted man falls in love with his artificially intelligent virtual assistant and the operating system seems to reciprocate his feelings.[14] In the television show *Black Mirror*, a grief-stricken widow purchases a program that imitates her recently deceased husband by learning from his text messages, emails, and online postings, and even mimics his voice from recordings.[15] Compared to these more ostentatious fictional representations, machine learning systems of the modern age are so elegantly intertwined with our daily lives that they are virtually undetectable. At the heart of all the various machine learning systems we encounter everyday are the algorithms that define their purpose and scope.

[13] DEP'T OF HOMELAND SEC., S&T ARTIFICIAL INTELLIGENCE & MACHINE LEARNING STRATEGIC PLAN 3 (Aug. 2021), https://www.dhs.gov/sites/default/files/publications/21_0730_st_ai_ml_strategic_plan_2021.pdf.

[14] HER (Annapurna Pictures 2013).

[15] *Black Mirror: Be Right Back* (Zeppotron Feb. 11, 2013).

A. *What Is an Algorithm?*

In its most basic form, an algorithm is "any well-defined computational procedure that takes some value, or set of values, as input and produces some value, or set of values, as output."[16] The series of steps taken to turn the input into the output is the algorithm. A non-mathematical example is a recipe. Using a recipe, an individual takes the input, the ingredients, and conducts a series of steps which transform the input into the output. Simple computational algorithms are not so different. In computer programming, an algorithm is a very specific set of instructions given to a computer program to complete a task.[17] Sorting algorithms, for example, can be used to take a series of values, called an array, and sort them in ascending order by following a list of clearly defined mathematical steps.[18] While most people can sort a short list of numbers very quickly without the use of a computer algorithm, this obviously becomes increasingly more time consuming the longer the list is. Algorithms are useful because humans, despite our ability to think and reason, are limited in our computational power and processing speed. Algorithms allow us to quickly distill large amounts of information and make connections that would otherwise go unnoticed in large data sets.

There are many ways to classify the various algorithms used in computer sciences, but the fundamental distinction for the purposes of this comment is between what we will call "traditional algorithms" and machine learning algorithms. Traditional algorithms are geared toward data organization and interpretation, whereas machine learning systems are most often used for their predictive capabilities.[19] Unlike the sorting algorithms discussed above, systems that rely on machine learning are not coded with the specific instructions necessary to complete a task. Rather, they are programed to "learn" how to complete the task on their own, through the use

[16] THOMAS H. CORMEN, CHARLES E. LEISERSON, RONALD L. RIVEST, & CLIFFORD STEIN, INTRODUCTION TO ALGORITHMS 5 (3d ed. 2009).

[17] An algorithm is a "prescribed set of well-defined rules or instructions for the solution of a problem, such as the performance of a calculation, in a finite number of steps." *Algorithm*, OXFORD REFERENCE: A DICTIONARY OF COMPUTER SCIENCE (Andrew Butterfield, Gerard Ekembe Ngondi, & Anne Kerr eds., 7th ed. 2016); Hannibal Travis, *Intelligent Entertainment: Shaping Policies on the Algorithmic Generation and Regulation of Creative Works*, 14 FIU L. REV. 179, 183 (2020) ("An algorithm is any stepwise solution to a problem or progress towards a goal . . . . A mathematical algorithm is a procedure for solving a mathematical problem.").

[18] CORMEN ET AL., *supra* note 16, at 16.

[19] Srinivas Rao, *How Does the ML Algorithm Differ from the Traditional Algorithm?,* MEDIUM (Mar. 17, 2020), https://medium.com/@raosrinivas2580/how-does-the-ml-algorithm-differ-from-the-traditional-algorithm-b7c3a2799e10.

of training data.[20] By providing a system with sufficient examples, the program can derive meaning, formulate rules, and make predictions.[21]

When it comes to algorithmic analysis of complex systems, whether it be shipping routes for international commerce or your own social media usage, machine learning programs use training data to create a model or digital mirror of whatever system it is analyzing that can be used to predict how the system will react in various situations and in response to different variables.[22] An algorithmic model operates like a sieve. Just as the programmed algorithm determines the objective of the system, a person constructs the sieve and defines its objective—if the objective is to strain flour, then the output is the flour that passes through, if the objective is to catch shark teeth in a riverbed, then the output is what is caught by the sieve. In a machine learning system, the programmers do not provide the program with the instructions on how to get to the stated result. Instead, the system gets there on its own, inching its way closer to the desired result through trial and error.[23] Rather than the sieve catching everything too large to pass through, it would be as if the sieve was eventually able to recognize and differentiate between what it is supposed to catch and what it is not supposed to catch.[24]

There are two main types of machine learning—supervised and unsupervised learning. In supervised machine learning, the training data is correctly labeled with whatever feature the programmer is trying to teach the program to recognize.[25] This is called a training set.[26] Programmers provide the system with examples of correctly classified data and using this labeled data, the program constructs a model against which it will compare new data to determine whether the new data fits within the parameters and rules it has derived from the training set.[27] Most people interact with this type of machine

[20] Solon Barocas et al., *Data & Civil Rights: Technology Primer*, DATA & SOC'Y RSCH. INST., at 4 (Oct. 30, 2014), http://www.datacivilrights.org/pubs/2014-1030/Technology.pdf.

[21] DEP'T OF HOMELAND SEC., *supra* note 13, at 2.

[22] O'NEIL, *supra* note 1, at 18 (defining a model as "an abstract representation of some process [that] takes what we know and uses it to predict responses in various situations").

[23] Barocas et al., *supra* note 20, at 4*; see* Janelle Shane, *The Danger of AI is Weirder than You Think*, TED (Apr. 2019), https://www.ted.com/talks/janelle_shane_the_danger_of_ai_is_weirder_than_you_think#t-114587.

[24] If this analogy seems strange and long-winded, note that machine learning is currently being used in the sorting of municipal waste and recycling whereby computers, through the use of training data, learn to recognize and sort recyclable materials. Carrissa Pahl, *How Machine Learning and Robotics Are Solving the Plastic Sorting Crisis*, PLUG & PLAY (Oct. 29, 2020), https://www.plugandplaytechcenter.com/resources/how-ai-and-robotics-are-solving-plastic-sorting-crisis/.

[25] Rich, *supra* note 1, at 881.

[26] *Id.*

[27] *Id.*

learning every day, as it is commonly used in data classification.[28] For example, using supervised machine learning, a computer program classifies incoming emails as "junk" and routes it to the appropriate folder.[29] Moreover, in systems that employ unsupervised machine learning, the program is not trained using a correctly classified training set.[30] Instead, the program is tasked with recognizing trends, formulating rules, and making predictions on its own.[31] Unsupervised learning is the type of machine learning used by many modern tech companies like Facebook and Google to personalize user experiences, maintain user engagement and target advertisements by learning from users' online behavior to predict future conduct.[32]

Despite how reliable a machine learning program may be, the term artificial intelligence is somewhat of a misnomer. These programs are highly efficient in analyzing large data sets, recognizing trends, and making predictions to reach a specified goal based on data, but they do not actually think the way humans do, or at all.[33] Machine learning systems are only optimized to analyze data to achieve or maintain the specifically preprogrammed result. The accuracy and predictive ability of machine learning programs is not based on the system's ability to think and reason but rather on the sheer quantity and quality of the training data.[34] Whereas a human brain may only need one data point to link a cause to a particular effect, a reliable statistical model will likely be based on "millions or billions of data points."[35] For example, using the billions of photos uploaded to Facebook, the company's algorithms can differentiate between a dog and a tree, or even individual people without understanding what a dog, tree, or person is.[36]

[28] *Cf. id.*

[29] *Id.* at 882.

[30] Emily Berman, *A Government of Laws and Not of Machines*, 98 B.U. L. REV. 1277, 1287 (2018).

[31] *Id.*

[32] Brent Barnhart, *Everything You Need to Know About Social Media Algorithms*, SPROUT SOC. (Mar. 26, 2021), https://sproutsocial.com/insights/social-media-algorithms/; O'NEIL, *supra* note 1, at 74–75.

[33] O'NEIL, *supra* note 1, at 75–76; Shane, *supra* note 23, at 1:50.

[34] O'NEIL, *supra* note 1, at 75–76.

[35] *Id.* at 76.

[36] Shane, *supra* note 23, at 1:57.

B. Applications of Machine Learning by Modern Tech Companies

Modern internet-based companies are utilizing supervised and unsupervised machine learning to collect and analyze immense quantities of data from billions of users for a multitude of different purposes.[37] Companies like Facebook, Google, Twitter, and Apple are just some examples of large companies that employ machine learning in their websites and applications. One of the most common applications of unsupervised machine learning is the personalization of online content using algorithms optimized to maintain or increase user engagement.

In the case of most machine learning systems optimized to personalize content, the algorithmic alchemy occurs during the decision-making process of the user. Based on the circumstances, an individual using the particular website or application makes a choice or takes some action; this action and the context of the action is logged by the system and then those choices become new input for the algorithm to analyze.[38] Eventually, by analyzing these choices and the context in which they were made and presenting more choices, the algorithm inches closer and closer to its defined objective which is generally to make the user spend as much time and money on the website as possible.[39] Eventually, the algorithm can predict behavior with a dazzling degree of accuracy. The use of machine learning to curate users' online experiences to their personal preferences is utilized by many different kinds of companies. Machine learning is at the heart of Netflix's recommendation algorithms that work to analyze an individual's watch history (among other factors) to recommend new movies or shows.[40] Amazon, too, uses these systems to suggest products to buy based on previous purchases.[41]

Social media companies are perhaps the most notorious for their use of machine learning to customize user experiences. The machine learning systems used by social media companies are usually optimized to maintain user engagement, increase the company's user base, and generate revenue. To do so, the machine learning systems collect user data, create a model of each user, and suggest content and advertisements that the algorithm predicts

[37] Morrison, *supra* note 5; O'NEIL, *supra* note 1, at 74–76.

[38] Allison Zakon, *Optimized for Addiction: Extending Product Liability Concepts to Defectively Designed Social Media Algorithms and Overcoming the Communications Decency Act*, 2020 WIS. L. REV. 1107, 1112–14 (2020).

[39] *Id.*

[40] *Machine Learning: Learning How to Entertain the World*, NETFLIX RSCH., https://research.netflix.com/research-area/machine-learning (last visited Apr. 27, 2023).

[41] Larry Hardesty, *The History of Amazon's Recommendation Algorithm*, AMAZON SCI. (Nov. 22, 2019), https://www.amazon.science/the-history-of-amazons-recommendation-algorithm.

the particular user is most likely to engage with.[42] Based on interactions with the site, the algorithm "learns" what content users engage with most and uses this information to make suggestions to tailor each user's experience.[43] The input data in this case is not limited to the information that users post publicly online such as photos and tweets, but is comprised of every action that the individual takes on the application or website.[44] In addition to the more concrete actions like posts, likes, comments, and search queries, the algorithms analyze more passive engagement data including the type of content the user is viewing, the time spent viewing particular content, the time the content is viewed, as well as the user's frequency of interactions with other users, all in relation to outside factors including location and time of year.[45]

Moreover, internet algorithms employed by some tech companies analyze all interactions that one has with the particular application or website and across platforms and websites owned by the parent company.[46] The largest tech companies that use these machine learning systems, like Google and Facebook, monitor user behavior across all associated platforms.[47] For example, Google's algorithms will collect and analyze browsing data on Chrome, watch history on YouTube, and location data in Maps. Facebook will catalogue every action taken on Facebook and Instagram.[48] Furthermore, in regard to Facebook, its algorithmic analysis is not restricted to the examination of user data on its own platforms. Rather, Facebook tracks users as well as non-users on third-party sites that employ Facebook's advertising and tracking tools.[49] Facebook Pixel is a tool that allows websites to track the activity of online shoppers while on their site to gain data on consumer habits and target advertisements.[50]

Another common application of supervised machine learning by tech companies is in photo classification. Facial recognition and identification technologies can learn to identify individual people from image databases—

[42] Zakon, *supra* note 38, at 1112–14.

[43] *Id.*

[44] *See* Adam Mosseri, *Shedding More Light on How Instagram Works*, INSTAGRAM (June 8, 2021), https://about.instagram.com/blog/announcements/shedding-more-light-on-how-instagram-works (explaining the types of behaviors logged and analyzed by Instagram's algorithms).

[45] *Id.*; Morrison, *supra* note 5; THE SOCIAL DILEMMA, at 16:40–17:20 (Exposure Labs, Argent Pictures, The Space Program 2020).

[46] Morrison, *supra* note 5.

[47] *Id.*

[48] *Id.*

[49] David Nield, *All the Ways Facebook Tracks You—and How to Limit It*, WIRED (Jan. 12, 2020, 7:00 AM), https://www.wired.com/story/ways-facebook-tracks-you-limit-it/.

[50] *Id.*

a practice that has become increasingly common in law enforcement.[51] For a decade, Facebook's DeepFace facial recognition algorithms have been used in conjunction with social media data to identify individuals posted in photos on the site.[52] This software was heavily criticized for its potentially wide-reaching surveillance applications.[53] In November 2021, Facebook announced its intention to delete the over one billion facial recognition templates that it has amassed over the years due to growing criticisms of the privacy implications and potential for misuse.[54] Apple also recently unveiled its new software which scans photos uploaded to iCloud for child sexual abuse material (CSAM). According to Apple, this technology would allow the company's algorithms to "flag accounts exceeding a threshold number of images that match a known database of CSAM image hashes so that Apple can provide relevant information to the National Center for Missing and Exploited Children."[55] Despite the potentially life-saving benefits of this technology, the announcement was incredibly controversial and was met with vehement protest from consumers as well as surveillance professionals for the program's potential for misuse.[56] Apple delayed the rollout of the feature to address these privacy concerns.[57]

Regardless of the particular variety of machine learning employed or task the system is optimized to perform, companies that utilize machine learning are all part of an upward trend of "data mining"—the accumulation of massive troves of internet data for large-scale algorithmic analysis to predict outcomes.[58] As discussed in the next section, the government also employs data mining and machine learning to aid in crime prevention and law enforcement and is often aided by private companies in these pursuits.

[51] Doktor, *supra* note 11, at 556–57; Kashmir Hill & Ryan Mac, *Facebook, Citing Societal Concerns, Plans to Shut Down Facial Recognition System,* N.Y. TIMES (Nov. 2, 2021, 12:27 AM), https://www.nytimes.com/2021/11/02/technology/facebook-facial-recognition.html.

[52] Hill & Mac, *supra* note 51.

[53] *Id*.

[54] *Id*.

[55] CSAM DETECTION: TECHNICAL SUMMARY, APPLE, at 3 (Aug. 2021), https://www.apple.com/child-safety/pdf/CSAM_Detection_Technical_Summary.pdf.

[56] Jon Porter, *Apple Scrubs Controversial CSAM Detection Feature from Webpage but Says Plans Haven't Changed*, THE VERGE (Dec. 15, 2021, 11:56 AM), https://www.theverge.com/2021/12/15/22837631/apple-csam-detection-child-safety-feature-webpage-removal-delay.

[57] *Id.*

[58] Caryn Devins, Teppo Felin, Stuart Kauffman, & Roger Koppl, *The Law and Big Data*, 27 CORNELL J. L. & PUB. POL'Y 357, 363–64 (2017).

C. Uses of Machine Learning and Digital Data by Law Enforcement

Just as many areas of crime have moved online, law enforcement, too, has expanded its investigations from traditional searches of physical property and effects to more surreptitious surveillance of digital data. Online phishing operations and child pornography are just some examples of internet-based crimes that leave cookie crumbs that could be detected through analysis of digital data and machine learning algorithms. The wealth of digital data available to law enforcement, coupled with the use of artificial intelligence, has bolstered police departments' investigative capabilities.[59] Law enforcement's newfound capacity for internet fact-finding and algorithmic data analysis has garnered praise, in light of its potential to assist in the capture of criminals, as well as criticisms for the potential privacy implications.[60] The government is limited in the amount of data it can collect on its own and because the success of predictive algorithms is contingent on the quantity and quality of input data, government entities have been looking to the private sector to assist in surveillance. Indeed, the "the government has essentially lost its monopoly on intelligence," and now relies heavily on data obtained by private companies.[61]

Machine learning is becoming increasingly popular in sentencing and policing.[62] In sentencing, personal data is incorporated into recidivism models to determine the likelihood that the individual will return to prison.[63] Predictive policing algorithms use crime data and associational information to predict areas that crime is more likely to occur.[64] By allowing a computer program to make these decisions, there is less of a chance of overt racism or bias by individual officers or judges. However, the use of algorithms in this capacity has also been criticized.[65] Given that algorithms do not think holistically, but rather interpret the data through the narrow lens of their specifically identified task, they fail to

[59] *See* Sara Morrison, *To Catch an Insurrectionist*, VOX (Jan. 6, 2022, 7:00 AM), https://www.vox.com/recode/22867000/january-6-fbi-search-facebook-google-insurrection.

[60] *Id.*

[61] *Big Data & Big Brother: The Rise of the Surveillance State and the Death of Privacy?*, HERITAGE FOUND., at 11:53–12:56 (Oct. 25, 2019, 12:00 PM), https://www.heritage.org/technology/event/big-data-big-brother-the-rise-the-surveillance-state-and-the-death-privacy.

[62] O'NEIL, *supra* note 1, at 23–27.

[63] *Id.*

[64] Rachel Levinson-Waldman, *Government Access to and Manipulation of Social Media: Legal and Policy Challenges,* 61 HOWARD L.J. 523, 554 (2018); O'NEIL, *supra* note 1, at 23–27.

[65] United States v. Curry, 965 F.3d 313, 344–45 (4th Cir. 2020) (Thacker, J., concurring) (discussing how predictive policing algorithms, though seemingly objective, are based on decades of crime data gained through racist policing); Renata M. O'Donnell, *Challenging Racist Predictive Policing Algorithms Under the Equal Protection Clause*, 94 N.Y.U. L. REV. 544, 547 (2019).

account for the bias already in the system.[66] Algorithms, by design, are reductive, and an algorithm's output is only as good as its input.[67] This concept is known as "garbage in, garbage out."[68] If the output is based on skewed data, the algorithm will not recognize the issue and correct it. Instead, the algorithm will just model the data and make predictions based on the model. For example, if a city's crime data is based on decades of over-policing of predominately Black neighborhoods, a predictive policing algorithm will draw the conclusion that crime is more likely to occur in that neighborhood, creating a feedback loop of disproportionate policing.[69] This also shows that while conclusions drawn by seemingly innocuous algorithms may be technically accurate, they often magnify issues hiding in the data.[70]

More and more, the government has been relying on data compiled by private companies, and law enforcement has increasingly been using data shared on social media sites to aid in their investigations.[71] Social media websites have become a tremendous source of evidence in modern criminal investigations.[72] This trend is especially evident in government use of facial recognition technology and social media data. One example is Clearview AI, a private artificial intelligence company that works in conjunction with hundreds of law enforcement agencies and police departments to provide databases of photos scraped from Facebook accounts.[73] The billions of photos are used with photo classification algorithms to identify suspects, a practice found to be illegal in many countries.[74] Furthermore, legal scholars counsel that social media data can be incorporated into predictive policing programs as a means of developing reasonable suspicion in determining when and where crimes are likely to occur and who is likely to carry them out.[75] Additionally, data scientists contemplate the use of machine learning to predict instances of civil

[66] *Curry,* 965 F.3d at 344–45; O'Donnell, *supra* note 65.

[67] David Lehr & Paul Ohm, *Playing with the Data: What Legal Scholars Should Learn About Machine Learning*, 51 U.C. DAVIS L. REV. 653, 701 (2017).

[68] *Id.* at 665.

[69] O'NEIL, *supra* note 1, at 23–27.

[70] *Id.*

[71] Justin P. Murphy & Adrian Fontecilla, *Social Media Evidence in Government Investigations and Criminal Proceedings: A Frontier of New Legal Issues*, 19 RICH. J.L. & TECH. 11, para. 10 (2013); *see* Morrison, *supra* note 59.

[72] *See* Murphy & Fontecilla, *supra* note 71; *see* Morrison, *supra* note 59.

[73] Andrew Guthrie Ferguson, *Facial Recognition and the Fourth Amendment*, 105 MINN. L. REV. 1105, 1121–22 (2021).

[74] Kashmir Hill, *Clearview AI's Facial Recognition App Called Illegal in Canada*, N.Y. TIMES (Feb. 3, 2021, 1:18 PM), https://www.nytimes.com/2021/02/03/technology/clearview-ai-illegal-canada.html.

[75] Levinson-Waldman, *supra* note 64, at 554.

unrest and insurrection.[76] Scientists believe that such technology can be utilized to predict events like the January 6th insurrection or potential coups.[77] To do so, machine learning systems analyze various types of data including economic trends, historical data, social media sentiment, as well as less obvious factors such as transportation disruptions and weather volatility.[78]

The use of digital data and machine learning in law enforcement has been the basis for much debate regarding privacy and the Fourth Amendment. The analysis of this new data under the Fourth Amendment requires a thorough analysis of the Supreme Court's Fourth Amendment jurisprudence and including the Court's treatment of digital data in recent cases.

III. THE FOURTH AMENDMENT AND DIGITAL DATA

The Fourth Amendment to the United States Constitution assures that "the right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated."[79] To determine whether the Fourth Amendment attaches to a particular government action, the Court distinguishes between actions that qualify as searches under the Fourth Amendment and those that do not.[80] In the centuries since its drafting, the logical framework with which the Court analyzes the Fourth Amendment has been expanded to make room for more varied types of property, and the Court has adopted a more expansive view of what government actions constitute a search in light of emerging technology.[81]

As the government increasingly relies on predictive algorithms and digital data collected by private companies to aid in its criminal investigations, courts and legal scholars have begun to grapple with the difficulties of applying traditional Fourth Amendment concepts to new forms of data and methods of data collection.[82] The Supreme Court has not yet considered algorithmic data in the context of the Fourth Amendment, but the

[76] Steven Zeitchik, *The Battle to Prevent Another Jan. 6 Features a New Weapon: The Algorithm*, WASH. POST (Jan. 6, 2022, 5:00 AM), https://www.washingtonpost.com/technology/2022/01/06/jan6-algorithms-prediction-violence/.

[77] *Id.*

[78] *Id.*

[79] U.S. CONST. amend. IV.

[80] *See* Doktor, *supra* note 11, at 559.

[81] *See, e.g.,* Kyllo v. United States, 533 U.S. 27, 40 (2001); Riley v. California, 573 U.S. 373, 401 (2014).

[82] *See* Patrick W. Nutter, *Machine Learning Evidence: Admissibility and Weight*, 21 U. PA. J. CONST. L. 919 (2019).

Court has addressed the difficulties of applying traditional Fourth Amendment rules and standards to varying forms of emerging technology.[83]

A. The Trespass Framework

From the initial drafting of the Fourth Amendment until the late 1960s, the Fourth Amendment was viewed solely in the context of common law trespass.[84] Conduct that was considered to be violative of this provision was confined to physical intrusions by the government upon an individual's person, house, papers, and effects.[85] Understandably, when the Founders drafted the Fourth Amendment it would have been difficult to imagine many kinds of intrusions that were not purely physical, and the Founders certainly could never have envisioned the surveillance capabilities of the modern age. Government conduct that today would constitute blatant and overt infringement on Fourth Amendment rights passed judicial scrutiny under the trespass rationale.[86]

In the 1928 case of *Olmstead v. United States*, the Supreme Court held that the act of wiretapping an individual's home phone did not constitute a search under the Fourth Amendment because the wiretapping was done without trespass onto the defendant's property.[87] Because the officers tapped the wires outside the curtilage of the defendant's home, the Court urged, there was no physical intrusion that could be considered a search under the Fourth Amendment.[88] Furthermore, the Court reasoned that there was no seizure because nothing tangible was taken by the government.[89] The Court stated that "the Amendment itself shows that the search is to be of material things" and the contents of a conversation transmitted over wires and intercepted by the government were not considered effects.[90]

In *Olmstead* and subsequent cases applying the trespass rationale to telephonic communication, the Court reasoned that by installing and using a telephone, one essentially "projects his voice" outside the confines of his home and therefore relinquishes any Fourth Amendment protections to that information.[91] This perspective, though now antiquated, was reasonable in

[83] *Riley*, 573 U.S. at 385.

[84] Katz v. United States, 389 U.S. 347, 352–53 (1967).

[85] Goldman v. United States, 316 U.S. 129, 134–36 (1942); Olmstead v. United States, 277 U.S. 438, 464–66 (1928).

[86] *Goldman*, 316 U.S. at 134–36; *Olmstead*, 277 U.S at 464–66.

[87] *Olmstead*, 277 U.S. at 457, 466.

[88] *Id.*

[89] *Id.* at 464–66.

[90] *Id.*

[91] *Id.* at 466; *Goldman*, 316 U.S. at 135.

light of the novelty of telephonic communication. *Olmstead* was decided just fifty years after the invention of the telephone, and at the time the case was decided only about thirty-five percent of homes had a telephone installed.[92] As telephones, and eventually wireless communication, have become keystones of modern communication, the Court's Fourth Amendment analysis expanded to provide increased protections in light of evolving technology.[93]

B. The Reasonable Expectation of Privacy Standard

In the 1967 case *Katz v. United States*, the Supreme Court revolutionized the Fourth Amendment framework by abandoning the common law trespass rationale in favor of the much more expansive reasonable expectation of privacy standard.[94] In *Katz*, the Court considered whether the government conducted a search under the Fourth Amendment when the FBI placed a wire-tap on the outside of the public phone booth where Katz placed a call.[95] The Court concluded that the FBI's electronic surveillance of the telephone booth constituted a search despite the fact that the government never physically trespassed upon Katz's property.[96] The Court famously noted, "the Fourth Amendment protects people, not places," and found that Katz was reasonable in assuming that his telephone conversation would remain private, and this expectation of privacy was not diminished by virtue of the fact that the conversation took place in a public booth.[97] Justice Harlan's concurrence laid out the two-part analysis for the current reasonable expectation of privacy standard.[98] The Fourth Amendment now protects against government invasions upon one's expectation of privacy that society has deemed reasonable.[99] Therefore, the violation of an individual's reasonable expectation of privacy is a search under the Fourth Amendment that generally requires a warrant supported by probable cause.

[92] *Percentage of Housing Units with Telephones in the United States from 1920 to 2008*, STATISTA (Sept. 30, 2010), https://www.statista.com/statistics/189959/housing-units-with-telephones-in-the-united-states-since-1920/.

[93] *See generally* Katz v. United States, 389 U.S. 347 (1967).

[94] *Id.* at 361 (Harlan, J., concurring).

[95] *Id.* at 348–49.

[96] *Id.* at 353.

[97] *Id.* at 351, 359.

[98] *Id.* at 361 (Harlan, J., concurring) ("My understanding of the rule that has emerged from prior decisions is that there is a twofold requirement, first that a person have exhibited an actual (subjective) expectation of privacy and, second, that the expectation be one that society is prepared to recognize as 'reasonable.'").

[99] *Id.*

Although the *Katz* ruling expanded the application of the Fourth Amendment, a subjective expectation of privacy in a place or activity is no guarantee of societal approbation, and the determination of what is reasonable has been the basis of much debate. Individuals generally do not have a reasonable expectation of privacy regarding information that they knowingly expose to the public or turn over to third parties.[100] In *United States v. Knotts*, for example, the Court noted that "[a] person traveling in an automobile on public thoroughfares has no reasonable expectation of privacy in his movements from one place to another."[101] If the government acquires this information from such a public display or disclosure to a third party, it is not considered a search and thus does not implicate the Fourth Amendment at all.[102]

Early Supreme Court jurisprudence on the third-party doctrine mainly involved the disclosure of information to government informants.[103] Divulging information to undercover informants, the Court reasoned, does not constitute a search under the Fourth Amendment.[104] In addition to information shared with other individuals, the Supreme Court has applied the third-party doctrine to information shared with business entities.[105] In *Smith v. Maryland* the Supreme Court considered whether the government violated the Fourth Amendment when officers had a suspected robber's telephone company place a pen register on their office phones to record the numbers dialed from the suspect's phone.[106] The Court held that this was not a search under the Fourth Amendment because the defendant had a reasonable expectation of privacy in the contents of his conversation and not in the numbers he called because it is common knowledge that telephone companies keep records of outgoing calls.[107] Similarly, in *United States v. Miller* the Court held that when an individual turns over records to a bank, those records are no longer protected under the Fourth Amendment.[108]

The existence of the third-party doctrine is essential to many criminal investigations, and the absence of the doctrine would severely limit law enforcement's investigative capabilities. However, critics of the third-party doctrine have argued that its application to information shared with business

[100] *See generally* Smith v. Maryland, 442 U.S. 735 (1979); United States v. Miller, 425 U.S. 435 (1976).

[101] United States v. Knotts, 460 U.S. 276, 281 (1983).

[102] *Id.* at 282–83.

[103] Orin S. Kerr, *The Case for the Third-Party Doctrine*, 107 MICH. L. REV. 561, 567 (2009).

[104] Lee v. United States, 343 U.S. 747, 753–54 (1952).

[105] Kerr, *supra* note 103, at 569.

[106] Smith v. Maryland, 442 U.S. 735, 736–37 (1979).

[107] *Id.* at 743–46.

[108] United States v. Miller, 425 U.S. 435, 436–37 (1976).

entities is at odds with an honest interpretation of the reasonable expectation of privacy standard.[109] Legal scholars note that conduct such as turning records over to a bank and making telephone calls are unavoidable tenants of modern life.[110] These critics argue that for the Court to suggest that anyone engaging in such activities assumes the risk of opening oneself up to government scrutiny is employing a willful blindness to the realities faced by most consumers.[111] Justices on the Court share in these sentiments. For example, in his dissenting opinion in *Smith v. Maryland,* Justice Marshall urged that "[i]t is idle to speak of 'assuming' risks in contexts where, as a practical matter, individuals have no realistic alternative."[112] Despite criticisms, the third-party doctrine has remained largely unchanged since its inception, with *Carpenter v. United States* being the first crack in an otherwise rigid framework.[113]

C. The Fourth Amendment and Emerging Technology

A series of recent cases highlight the Supreme Court's mounting Fourth Amendment jurisprudence regarding emerging technology that could not have been contemplated by the Founders. In the 2001 case *Kyllo v. United States*, the Court considered law enforcement use of thermal imaging technology on a suspect's home.[114] The Court found that use of sense-enhancing technology that is not widely available to the general public to investigate the confines of one's home is a search under the Fourth Amendment.[115] Justice Scalia, writing for the majority, reasoned that homeowners clearly have a reasonable expectation of privacy in their own homes, and to allow the government latitude to use technology to search what would otherwise be unsearchable would leave individuals "at the mercy of advancing technology."[116]

Furthermore, in 2012 the Court considered the extended use of a GPS tracking device on a suspect's car, and found that the use of such a device to track a suspect's movements for twenty-eight days was a search under the Fourth Amendment despite the fact that much of the suspect's movements

[109] Kerr, *supra* note 103, at 570–71; Christopher Slobogin & Joseph E. Schumacher, *Reasonable Expectations of Privacy and Autonomy in Fourth Amendment Cases: An Empirical Look at "Understandings Recognized and Permitted by Society,"* 42 DUKE L.J. 727, 732 (1993).

[110] Kerr, *supra* note 103, at 570–71.

[111] *Id.* at 571.

[112] *Smith*, 442 U.S. at 750 (Marshall, J., dissenting).

[113] *See generally* Carpenter v. United States, 138 S. Ct. 2206 (2018).

[114] Kyllo v. United States, 533 U.S. 27, 29 (2001).

[115] *Id.* at 40.

[116] *Id.* at 34–35.

occurred in public spaces.[117] In addition to the apparent departure from the traditional knowing exposure rationale detailed in *Knotts*, the Court relied on the all-but-forgotten trespass framework in making their decision.[118] The Court reasoned that the government physically trespassed upon the suspect's property when they used the GPS device to track his movements for such an extended period of time.[119] Concurring in the judgment, Justice Sotomayor discussed the potential effects of modern government surveillance techniques on the *Katz* reasonable expectation of privacy analysis.[120] Justice Sotomayor questioned whether a reasonable person would expect that the entirety of their movements would be aggregated in such "a manner that enables the government to ascertain . . . their political and religious beliefs, sexual habits, and so on." She further considered the propriety of the third-party doctrine regarding digital data, noting that many individuals reveal significant amounts of personal information to third parties while "carrying out mundane tasks."[121]

In *Riley v. United States*, the Court echoed much of Justice Sotomayor's sentiments and again reiterated the distinction between digital data and traditional physical evidence in holding the officers must secure a warrant before searching an individual's cell phone incident to an arrest.[122] Justice Roberts, writing for a unanimous Court, explained that the massive quantity of intimate content that people tend to store on smartphones can be used to create "a revealing montage of the user's life" that could be far more revealing than a search of one's home.[123] In carrying a smartphone, the Court noted, each person carries with them "a digital record of nearly every aspect of their lives—from the mundane to the intimate."[124] The Court further distinguished digital data stored on smartphones based on the pervasiveness of cell phone usage, noting that over ninety percent of Americans own a cell phone—a practice that is so pervasive in modern society that aliens from a distant planet would assume that these devices are part of our anatomy.[125]

[117] United States v. Jones, 565 U.S. 400, 402–03 (2012).

[118] *Id.* at 404–05.

[119] *Id.*

[120] *Id.* at 415 (Sotomayor, J., concurring).

[121] *Id.* at 416–17.

[122] Riley v. California, 573 U.S. 373, 401–03 (2014).

[123] *Id.* at 396.

[124] *Id.* at 395.

[125] *Id.* at 385.

D. Carpenter v. United States and Its Progeny

In 2018, the Supreme Court identified a narrow exception to the third-party doctrine in regards to certain digital data.[126] The Court determined that the fact that information being sought is in the possession of a third-party is not necessarily dispositive.[127] Timothy Carpenter was suspected of committing a series of robberies, and the government ordered Carpenter's cell phone company to turn over his cell-site location information (CSLI) for a period of 152 days.[128] Cell phones constantly scan their environments searching for the strongest signal from nearby cell towers, and whenever a cellphone connects to a cell site a time-stamped record is automatically logged by the cell phone company.[129] This happens several times a minute, and these records are often kept for years.[130] Using Carpenter's CSLI, the government was able to confirm that Carpenter was near the locations of the robberies when they occurred. The Sixth Circuit Court of Appeals held that "Carpenter lacked a reasonable expectation of privacy in the location information collected by the FBI because he had shared that information with his wireless carriers."[131] On certiorari, the Court found that the third-party doctrine does not apply to CSLI, so the government needed a warrant supported by probable cause to obtain Carpenter's CSLI.[132] A court order supported only by reasonable suspicion that the records were relevant to the ongoing investigation was therefore insufficient.[133]

In coming to this conclusion, the Court acknowledged a set of characteristics that separates CSLI from other types of information turned over to third parties like the bank and telephone records from early third-party doctrine jurisprudence. The Court noted that this information deserves increased protection "in light of the deeply revealing nature of CSLI, its depth, breadth, and comprehensive reach, and the inescapable and automatic nature of its collection."[134] The Court reasoned that using CSLI the government can gain tremendous insight into the intimacies of a person's life, "revealing not only his particular movements but through them his 'familial,

[126] *See generally* Carpenter v. United States, 138 S. Ct. 2206 (2018).

[127] *Id.* at 2217 (finding that "the fact that the information is held by a third-party does not by itself overcome the users claim to fourth amendment protection").

[128] *Id.* at 2212.

[129] *Id.* at 2211–12.

[130] *Id.*

[131] *Id.* at 2213.

[132] *Id.* at 2221.

[133] *Id.*

[134] *Id.* at 2223.

political, professional, religious, and sexual associations.'"[135] Moreover, the Court differentiated the practice of accessing CSLI from other types of government surveillance. Because of the prevalence of cell phones in modern society, the Court reasoned that "this newfound tracking capacity runs against everyone."[136] Therefore, with access to this data, the government need not surveil any particular person. Rather, it is as if everyone who owns a cell phone is constantly being surveilled regardless of whether they are suspected of a crime and the government need only tap into this database should the suspicion occur.

Although the holding in *Carpenter* only applied to CSLI, federal courts have already begun applying *Carpenter*'s rationale to cases involving information beyond cell site location information, including information posted to both private and public social media sites.[137] A North Carolina court held that a defendant had a reasonable expectation of privacy regarding information he posted on a private Facebook page, therefore the government needed a warrant to search his page.[138] The court found that the third-party doctrine did not apply to the information at issue because the non-public Facebook posts and messages were not directed to Facebook.[139] In this case Facebook was only an intermediary through which the defendant sent the information to the intended recipients.[140] Additionally, a Kansas federal court extended this reasoning to information posted on a public account.[141] They found that the defendant had a reasonable expectation of privacy in information posted on a public Facebook page so a government search of such information with a facially invalid, overbroad warrant was an unreasonable search under the Fourth Amendment.[142]

IV. MACHINE LEARNING DATA UNDER *CARPENTER V. UNITED STATES*

Government access to and manipulation of user engagement data compiled by machine learning algorithms sparks many of the same privacy concerns detailed by the Court throughout its recent Fourth Amendment

[135] *Id.* at 2217 (quoting United States v. Jones, 565 U.S. 400, 415 (2012) (Sotomayor, J., concurring)).

[136] *Id.* at 2218.

[137] *See* United States v. Chavez, 423 F. Supp. 3d 194, 204 (W.D.N.C. 2019); United States v. Irving, 347 F. Supp. 3d 615, 621 (D. Kan. 2018).

[138] *Chavez*, 423 F. Supp. 3d at 204.

[139] *Id.*

[140] *Id.*

[141] *Irving*, 347 F. Supp. 3d at 621.

[142] *Id.* at 625.

jurisprudence on digital data. This Section contemplates the extension of *Carpenter* to internet engagement data accumulated by private tech companies through machine learning. The *Carpenter* Court found that CSLI was distinct due to the intimate and revealing nature of the information collected, the automatic and continuous nature of the data collection, and the fact that cell phone usage is an incredibly prevalent and insistent part of daily life.[143] As will be discussed below, machine learning data fits neatly under this rationale.

A. The Intimate and Revealing Nature of the Data

The Court in *Carpenter* reiterated much of the Supreme Court's recent sentiments regarding the intimate nature of digital data in finding that CSLI is wholly different than other types of personal information.[144] In recent holdings, the Court has distinguished digital information that reveals intimate details of an individual's life including their "familial, political, professional, religious, and sexual associations."[145] Compared to the call logs gathered from pen registries in *Smith v. Maryland* that are limited in their probative value, the *Carpenter* Court reasoned that, using CSLI, the government can paint a stunningly accurate portrait of a suspect's life.[146] Likewise, through the algorithmic analysis of the minutiae of an individual's online engagement machine learning systems construct detailed digital models of each user.[147] Experts in the field counsel that "every single action you take [on social media sites] is carefully monitored and recorded" and that "[networking sites like Facebook and Instagram] have more information about [people] than has ever been imagined in human history."[148] The various machine learning algorithms used by networking sites collect and analyze biometric data from posted photos; catalogue associations and frequency correspondence between users; record the amount of time spent engaging with particular content; and analyze each click, scroll, like, comment, and search query as a means of personalizing content and targeting advertisements.[149] Tristan Harris, founder of the Center for Humane Technology, counsels that, armed with this quantity of data, social media algorithms are so efficient at analyzing individual characteristics that they are able to determine when

[143] Carpenter v. United States, 138 S. Ct. 2217, 2220, 2223 (2018).

[144] *Id.* at 2217.

[145] Riley v. California, 573 U.S. 373, 396 (2014) (quoting United States v. Jones, 565 U.S. 400, 415 (2012) (Sotomayor, J., concurring)).

[146] *Carpenter*, 138 S. Ct. at 2217–19.

[147] Morrison, *supra* note 5; Zakon, *supra* note 38, at 1112–14.

[148] THE SOCIAL DILEMMA, *supra* note 45.

[149] *Id.*; Morrison, *supra* note 5.

people are lonely or depressed, can detect personality traits like introversion or extroversion, and can classify individuals based on the particular neurosis they may have.[150]

Furthermore, the Courts in *Carpenter* and *Jones* acknowledged potential privacy concerns regarding long-term data aggregation.[151] In discussing the probative value of digital data, the Justices on the Court, as well as legal scholars, have noted that although the individual bits of information might not be revealing in nature, in the aggregate this data has significant probative worth.[152] Justice Sotomayor, in questioning the application of the third-party doctrine to digital data, argued that people often convey incredibly personal information while carrying out mundane tasks.[153] She urged that most individuals would not reasonably assume that in engaging in the mundanity of their lives they turn over data that could be aggregated to evidence intimate characteristics like religious beliefs or sexual habits.[154] In regard to the engagement data gathered by machine learning systems, while each individual click, scroll, or like might seem inconsequential, in the aggregate these millions of data points create a detailed digital record of an individual's online habits, evidencing one's likes, dislikes, associations, and personality traits. This is precisely the category of information that the Court has urged is distinct and worthy of increased protections.

This type of intimate and revealing data can have tremendous probative value for law enforcement agencies. Social media data is already used in conjunction with machine learning and legal scholars have discussed the potential application of algorithmic analysis to social media data as a means of developing reasonable suspicion.[155] Given the deeply revealing nature of the information collected by these machine learning systems and the current lack of constitutional protections in place regarding this data, the Court should take its first opportunity to address machine learning engagement data in the context of the Fourth Amendment.

B. Automatic and Continuous Data Collection

The Court further distinguished CSLI from other types of information based on the fact that CSLI is not truly shared with third parties but rather is collected automatically and continuously without the user's overt consent.[156]

[150] THE SOCIAL DILEMMA, *supra* note 45.

[151] Ferguson, *supra* note 73, at 1134–35.

[152] *Id.*; United States v. Jones, 565 U.S. 400, 417 (2012) (Sotomayor, J., concurring).

[153] *Jones*, 565 U.S. at 417 (Sotomayor, J., concurring).

[154] *Id.*

[155] Rich, *supra* note 1, at 873–74.

[156] Carpenter v. United States, 138 S. Ct. 2206, 2220, 2223 (2018).

Likewise, the user engagement data collected using machine learning algorithms is not voluntarily conveyed as has been traditionally understood by the third-party doctrine. Whereas public posts on social media may be more likely to be considered "shared" under the Doctrine, the constant and continuous collection of user engagement data is "not truly 'shared' as one normally understands the term."[157] Compared to an explicit act of divulging an incriminating secret to a government informant, for example, it is unlikely that many people would think that they were giving something to anyone by engaging with Facebook or Google applications by viewing content or liking posts. However, in engaging with social media sites as well as applications like Google and even Netflix, users automatically give engagement data to companies that is used to predict personal preference and conduct. As Justice Sotomayor explained in *Jones*, it is unlikely that an individual would reasonably expect that such probative content could be gleaned from the continuous analysis of such ordinary conduct.[158]

The Court in *Carpenter* distinguished bank and telephone records from CSLI and found that when *Smith* and *Miller* were decided, no one could have imagined that each individual would carry with them the technology that would allow the government to track the entirety of their movements so continuously.[159] Similarly, unlike other business records, the aggregation and use of engagement data by machine learning systems is a continuous and automatic form of data collection that would have been inconceivable when these cases were decided. Various networking algorithms from a variety of web-based applications constantly collect, store, and analyze personal data. The Court in *Carpenter* further highlighted the "inescapable" nature of CSLI data collection.[160] In regard to machine learning data, if one wishes to escape the automatic and continuous collection of engagement data, one would essentially have to stop using the internet altogether, a practical impossibility given the degree to which we all rely on the internet in our daily lives.

C. Prevalence in Modern Society

Another theme that is often discussed by the Court when analyzing digital data in the context of the Fourth Amendment is the pervasiveness of the technology throughout society such that it has become practically unavoidable to expose oneself to government scrutiny.[161] The Court in

[157] *Id.* at 2220.

[158] *Jones*, 565 U.S. at 417.

[159] *Carpenter*, 138 S. Ct. at 2216–17.

[160] *Id.* at 2223.

[161] *See, e.g.*, *id.* at 2220; Riley v. California, 573 U.S. 373, 395 (2014).

Carpenter reasoned that CSLI is distinct based on the pervasiveness of cell phone usage in modern society which renders it an insistent part of daily life.[162] To allow the government to access this data without the protections of the Fourth Amendment, urged the Court, would expose the majority of the population to government surveillance simply because they engaged in a practice that is all but necessary for participation in modern life.[163] Likewise, as technology has proliferated, the act of engaging with websites and applications that collect user engagement data through machine learning algorithms has become less of a luxury and more so a "personal or professional necessity."[164]

Since the conception of the modern third-party doctrine, Justices on the Court have questioned the propriety of the doctrine as applied to information that is all but necessary to turn over to participate in society.[165] Justice Marshall, in his dissenting opinion in *Smith v. Maryland*, argued that one does not assume the risk of exposing oneself to government scrutiny by using a telephone because "[i]mplicit in the concept of assumption of risk is some notion of choice."[166] This element of pervasiveness was also present in *Riley* in regard to digital data stored on cell phones.[167] Chief Justice Roberts noted that the use of cellphones is such an insistent part of modern life that the person who does not carry with them a trove of personal data everywhere they go is the exception.[168] The *Carpenter* Court echoed this rationale in urging that using a cell phone and creating a log of CSLI is "indispensable to participation in modern society."[169]

Similarly, it is all but impossible to participate in society today without providing tech companies with millions of data points to filter through their user engagement and recommendation algorithms. Facebook, Instagram, Snapchat, Twitter, Google, YouTube, Netflix, and Amazon are just some of the common websites and applications that employ algorithmic analysis of user engagement data to personalize content. The pervasiveness of algorithmic analysis of engagement data is even more striking when one considers the fact that companies analyze data across their various platforms and even on third-party websites that employ their tracking and advertising tools.[170]

[162] *Carpenter*, 138 S. Ct. at 2220.

[163] *Id.*

[164] Smith v. Maryland, 442 U.S. 735, 749–50 (1979) (Marshall, J., dissenting).

[165] *Id.*

[166] *Id.*

[167] Riley v. California, 573 U.S. 373, 395 (2014).

[168] *Id.*

[169] Carpenter v. United States, 138 S. Ct. 2206, 2220 (2018).

[170] Nield, *supra* note 49.

Furthermore, the use of social media websites and applications, though not yet so saturated in society as cell phone usage, has proliferated greatly in recent years. The majority of people in the United States now have at least some form of social media.[171] In 2021, eighty-two percent of the population reported that they had a social networking profile, and this number is expected to rise.[172] Furthermore, the collective time spent on social media sites amongst Americans further speaks to the pervasive quality of social media usage in society. In 2020, the average American spent an average of 1,300 hours on social media, and the average Gen Z American spent an average of nine hours daily consuming screen media.[173] All this time spent engaging with social media content provides tech companies with the millions of data points, which they use to construct their models to personalize content. If this upward trend of social media use and digital data consumption continues, the act of providing companies with engagement data will certainly be less of a choice and more of a practical necessity to participate in modern life. This implicates the same concerns as CSLI because at that point the government need not explicitly surveil any suspect but rather just seek to obtain the information from the third-party that has been collecting it.

Some argue that if one truly does not wish to be subject to machine learning surveillance, one need not engage with any products that employ this practice because the use of these companies' products and applications is still a mere luxury.[174] However, this argument rests on a foundation that likely will no longer exist in the coming decades. Social networking sites, for example, have become more than just online platforms to share photos or opinions. Applications like Facebook, Instagram, and Twitter have become substantial sources of news and information, not just from other users but from official, reputable news sources and government officials.[175] In fact, over half of Americans report using social media as a news source.[176] As we increasingly spend more time engaging with websites and applications that

[171] S. Dixon, *Share of U.S. Population Who Use Social Media 2008–2021*, STATISTA (July 27, 2022), https://www.statista.com/statistics/273476/percentage-of-us-population-with-a-social-network-profile/.

[172] *Id.*

[173] Peter Suciu, *Americans Spent on Average More than 1,300 Hours on Social Media Last Year*, FORBES (June 24, 2021, 3:47 PM), https://www.forbes.com/sites/petersuciu/2021/06/24/americans-spent-more-than-1300-hours-on-social-media/?sh=4291656e2547.

[174] *Big Data & Big Brother: The Rise of the Surveillance State and the Death of Privacy?*, *supra* note 61, at 23:45–24:20, 26:45–28:10 (arguing that every person can avoid this type of surveillance because every consumer can choose whether or not they use Google and Facebook products, for example).

[175] Elisa Shearer, *More than Eight-in-Ten Americans Get News from Digital Devices*, PEW RSCH. CTR. (Jan. 12, 2021), https://www.pewresearch.org/fact-tank/2021/01/12/more-than-eight-in-ten-americans-get-news-from-digital-devices/.

[176] *Id.*

employ a large scale algorithmic analysis of our engagement data for leisure, work, school, and as a source of connectivity and news, the Court must consider the privacy protections that should be afforded to this new breed of personal data.

V. CONCLUSION

The same qualities that render the third-party doctrine inapplicable to CSLI are present in user engagement data collected by machine learning systems. The fact that private surveillance and machine learning has made it easier to surveil a population and catch criminals does not relieve people of the rights explicitly guaranteed to them. The Supreme Court in *Carpenter* assured that their ruling was narrow, but as technology continues to proliferate and machine learning becomes more pervasive and accurate, it will be difficult for the Court to ignore the incredible surveillance potential of large corporations armed with machine learning and massive quantities of personal data. At the very least, the Supreme Court should take its first opportunity to address the Fourth Amendment and privacy concerns implicated in the government's use of large quantities of data aggregated by private companies. Technologies, like those offered by Facebook, Apple, and Google, have the remarkable potential to simplify, inform, entertain, and unify. We must ask whether the machine learning technology and algorithmic analysis of personal data that works to aid in these pursuits has a place in law enforcement, and we must ensure that in reaping the benefits of the digital age we do not run afoul of the liberty interests upon which our nation was founded.