THE IMPLICATIONS OF DATA SCRAPING: IT BENEFITS BIG BUSINESS, BUT WHAT DOES IT MEAN FOR YOU?

CASEY REILLY*

I. Introduction

The expansion of the digital world has drastically altered our everyday lives – from the way we order our morning coffee to the commencement of a new job search – many tasks occur at the tap of a finger.1 Today, consumers and businesses often interact through the intermediary of a mobile application or internet webpage.2 Studies

---

1 See Digital Around the World, DATAREPORTAL (Oct. 19, 2022), archived at https://perma.cc/66NV-M44S (finding that 65.7 percent of the world’s total population use the internet and average users spend 6 hours and 41 minutes online per day). “[M]obile phones now account for roughly 57 percent of our online time, as well as close to 53 percent of the world’s web traffic.” Id. See also Mobile App Download Statistics & Usage Statistics (2022), BUILDFIRE (Oct. 19, 2022), archived at https://perma.cc/E4M8-8LR7 (providing that “the average American checks their phone 262 times per day . . . [t]hat’s once every 5.5 minutes.”). “88% of mobile time is spent on apps.” Id.

2 See Scott Rigdon, How Has the Internet Changed Communication in Business?, VONAGE (Oct. 12, 2022), archived at https://perma.cc/N44R-N9EV (explaining impact of internet in business communications). Businesses used to be able to put ads in a local paper and cater to the needs of local customers, but this drastically changed with the inception of the internet. Id. Now, a business doesn’t have to rely on its locality to survive; there is a worldwide audience through the internet that has changed the way a business communicates, finds employees, and manages its competition. Id. See also Andrew Meola, Mobile apps are growing rapidly in this one area, INSIDER (May 12, 2016), archived at https://perma.cc/9XTK-MQL9 (stating that in 2016 mobile order-ahead apps represented a small share of consumer interaction but will be a $38 billion industry by 2020); Bindu Yennamsetti, Why Mobile Apps are Important for your Business, MySTORY (Feb. 20, 2020), archived at https://perma.cc/Y3FN-QV6U (describing the benefits of mobile application use,
show that the average person has over 80 applications on their smartphone; behind each and every one of these are respective companies, accounts, user agreements, and most importantly, copious amounts of data.\textsuperscript{3} This data has become a high-valued asset that ultimately forces companies to re-adjust their business models or risk losing their competitiveness in the market.\textsuperscript{4}

including increase of brand presence, customer loyalty, improved accessibility, and customer service).

\textsuperscript{3} See Mobile App Download Statistics & Usage Statistics (2022), supra note 1 (finding that the average smartphone has over 80 apps installed). “The average person uses 9 mobile apps per day and 30 apps per month.” \textit{Id}. As of 2020, Facebook, Instagram, YouTube, Gmail, and Twitter were among the top apps that users said they “can’t live without”. \textit{Id}. See also The 2022 Social Media Demographics Guide, KHOROS (2022), archived at https://perma.cc/88R3-7HAJ (providing number of users and demographics for various social media applications). As of 2022, Facebook has roughly 2.7 billion monthly active users, while Instagram has 1 billion monthly active users and 500 million active daily users. \textit{Id}. Meanwhile, LinkedIn, the primary professional networking site, has roughly 260 million monthly active users. \textit{Id}. See also Jessica Guynn, \textit{What you need to know before clicking ‘I agree’ on that terms of service agreement or privacy policy}, USA TODAY (Jan. 28, 2020), archived at https://perma.cc/MQ52-FY6R (reiterating a Deloitte survey, which found that “91% of people consent to terms of service without reading them. For younger people, ages 18-34, that rate was even higher: 97% did so.”). Terms of service “protects the company and explains to consumers what the rules are when using the service[].” \textit{Id}. A privacy policy “explains to users how their data will be collected and used by the company and any third parties or affiliates.” \textit{Id}. When a consumer clicks “I agree” on these documents, their approval becomes legally binding. \textit{Id}. A lot of what is in these documents is boilerplate, but granting a company the right to sell your personal information to third parties may be an area to pay extra attention to. \textit{Id}.

\textsuperscript{4} See Pauline Glikman & Nicolas Glady, \textit{What’s The Value Of Your Data?}, TECHCRUNCH (Oct. 13, 2015), archived at https://perma.cc/TAE3-47FY (categorizing data as a “strategic asset that allows companies to acquire or maintain a competitive edge.”). “For the shareholder, data embodies a financial potential. For the company itself, data can be used to optimize the way it does business . . . the value of personal data still depends largely on who uses it, how it is being used and in what context.” \textit{Id}. See also Nick Rea & Adam Sutton, \textit{PUTTING A VALUE ON DATA} (PwC, 2019) (stating that “organisations are beginning to understand the value of being viewed as data-centric[].”). Valuation multiples for data-driven firms tend to be significantly higher than those of their peers or other industries. \textit{Id}. See also Timothy Morey et al., \textit{Customer Data: Designing for Transparency and Trust}, HARV. BUS. REV. (May 2015), archived at https://perma.cc/TS25-5V5H (describing the expanding scope of data). “[I]ntelligent technology in physical products has allowed companies in many industries to collect new types of information, including users’ locations and behavior. The personalization this data allows, such as constant adaptation to users’ preferences, has become central to the product experience.” \textit{Id}. 

Organizations collect data in numerous ways for diversified purposes.\(^5\) Two methods of data retrieval, referred to as web scraping and data scraping, have become quite prevalent among business practices.\(^6\) Web scraping is the process of using an application to extract data from a website, while data scraping does not require the internet and simply refers to the extraction of data from one output-generated program to another.\(^7\) Data scraping offers extraordinary...
benefits for companies; businesses leverage data to enhance their brand strategy, improve marketing, increase sales, and remain competitive in pricing. Some companies have even built their entire business model around the technique of scraping. At the heart of this data’s value is the personal information of everyday people, like you

Data scraping is “when you take any publicly available data, whether it is on the web or your computer, and import the found information into any local file on your computer.” Id. Because it does not require the internet to be conducted, it is an extremely effective data retrieval method. Id. On the other hand, web scraping is “when you take any publicly available online data and import the found information into any local file on your computer.” Id. 8

See Fatenaitė, supra note 7 (referencing the power data scraping can have on organizations and it’s tremendous influence on competitor analysis and pricing, marketing and sales, product development, public relations, brand management, and strategy.); What is data scraping?, supra note 7 (describing the many purposes scraper bots can be designed for). “Content scraping – content can be pulled from the website in order to . . . replicate the unique advantage of a particular product or service that relies on content.” Id. With price scraping, “competitors are able to aggregate information about their competition.” Id. With contact scraping, “a scraper is able to aggregate contact details for bulk mailing lists, robocalls, or malicious social engineering attempts.” Id. See also Venky Anant et al., The consumer-data opportunity and the privacy imperative, MCKINSEY & CO. (Apr. 27, 2020) archived at https://perma.cc/7Z8K-AFXB (explaining that data, including location tracking, or other kinds of “personally identifiable information, are immensely valuable to companies: many organizations, for example, use data to better understand the consumer’s pain points and unmet needs. These insights help to develop new products and services, as well as to personalize advertising and marketing . . . ”).

9 See Freedman, supra note 5 (acknowledging “[s]ome companies have built an entire business model around consumer data, whether they sell personal information to a third party or create targeted ads to promote their products and services.”). See also Wendy Davis, Privacy Group Sides Against Scraper In Battle Over LinkedIn Users’ Data, DIGITAL NEWS DAILY (Oct. 20, 2017), archived at https://perma.cc/82SC-PHKY (outlining that “HiQ scrapes LinkedIn’s publicly available pages, analyzes the information to determine which employees are at risk of being poached, and then sells its findings to employers.”); Hossein Rahnama & Alex “Sandy” Pentland, The New Rules of Data Privacy, HARV. BUS. REV. (Feb. 25, 2022), archived at https://perma.cc/42LP-T2BQ (stating personal data is the “wellspring for millions of small businesses and countless startups, which turn it into customer insights, market predictions, and personalized digital services.”).
and I.\textsuperscript{10} What level of say do we have in the way our “assets” are being utilized as assets among big business?\textsuperscript{11}

This note addresses the implications of upholding data scraping as legal within the context of the business world and in relation to consumers’ personal information. This note begins by discussing the Computer Fraud and Abuse Act ("CFAA"), where the legality behind data scraping derives, and proceeds to explore the variations of cyber security and data privacy regulations at the federal and state levels. As the valuation of consumer data exponentially increases for companies, the methods they employ to protect this data are more important than ever before. This note argues that amendments to the CFAA and nationwide privacy protection laws are critical in ensuring that companies’ internal procedures are securely maintaining consumers’ personal information. Additionally, this note proposes amendments to the CFAA and suggests that an overarching federal data privacy regulation is critical for consumer protection.

\textsuperscript{10} See Freedman, supra note 5 (explaining the four categories of consumer data businesses collect). Personal data includes “personally identifiable information such as Social Security numbers and gender, as well as nonpersonally identifiable information, including your IP address, web browser cookies and device IDs . . . .” \textit{Id}. Engagement data “details how consumers interact with a business’s website, mobile apps, text messages, social media pages . . . .” and more. \textit{Id}. Behavioral data includes “transactional details such as purchase histories, product usage information . . . . and qualitative data . . . .” \textit{Id}. Attitudinal data “encompasses metrics on consumer satisfaction, purchase criteria, [and] product desirability . . . .” \textit{Id}.

\textsuperscript{11} See Rahnama & Pentland, supra note 9 (describing the change in handling consumer data).

Data was considered company property and a proprietary secret, even though the data originated from customers’ private behavior. That curtain has since been lifted and a convergence of consumer, government, and market forces are now giving users more control over the data they generate. Instead of serving as a resource that can be freely harvested, countries in every region of the world have begun to treat personal data as an asset owned by individuals and held in trust by firms. \textit{Id}. 
II. History

A. The History of the Computer Fraud and Abuse Act ("CFAA")

In 1984, Congress enacted the Comprehensive Crime Control Act ("CCCA"), which included the first federal computer crime statute, the Counterfeit Access Device and Computer Fraud and Abuse Act ("CADCFAA").12 This act was relatively narrow in scope, as it prohibited three specific types of conduct: unauthorized computer access to obtain security information, unauthorized computer access to obtain financial information, and trespass of a government computer.13 Two years later, Congress expanded the computer crime statute via 18 U.S.C § 1030, or what is commonly known today as the CFAA, to criminalize knowingly accessing a computer without authorization or


13 See Berris, supra note 12, at 3 (describing the CADCFAA as prohibiting three types of conduct, “(1) obtaining national security information through unauthorized computer access; (2) obtaining financial information through unauthorized computer access, and (3) trespassing into a government computer and ‘knowingly us[ing], modif[y]ing], destroy[ing], or disclos[ing] information’ on that computer.”). See also Din, supra note 12, at 407 (stating the nature of the statute as quite general). Although these phrases indicate that Congress was beginning to understand the nature of computer hacking, the phrases’ generality demonstrates that Congress only had a rudimentary framework to define this new type of crime. . . . the CADCFAA only prohibited the hacking of certain types of information, such as matters concerning national security, foreign relations, and financial credit. It also only applied to select computers, such as those that were operated for or on behalf of the government or those that belonged to financial institutions and contained financial records.

Id.
exceeding authorized access.\textsuperscript{14} The 1986 amendments prohibited additional types of conduct, which included modifying the intent requirements under the statute, prohibiting damage of unauthorized access to computers, and password trafficking.\textsuperscript{15}

In 1994, Congress added a civil cause of action to the CFAA that allowed private parties to obtain compensatory damage, injunctive, or equitable relief.\textsuperscript{16} These amendments further broadened

\textsuperscript{14} See 18 U.S.C. § 1030 (providing codification of CFAA). See also CFAA Background, supra note 12 (describing that congress expanded “the computer crime statute by passing the Computer Fraud and Abuse Act (‘CFAA’) . . . [t]echnically speaking, the CFAA was the 1986 amendment to 18 U.S.C. § 1030; however, 18 U.S.C. § 1030 in its entirety is commonly referred to as the Computer Fraud and Abuse Act and vice versa.”); Berris, supra note 12 (summarizing the CFAA). “The Computer Fraud and Abuse Act (CFAA), 18 U.S.C. § 1030, is a civil and criminal cybercrime law prohibiting a variety of computer related conduct. Although sometimes described as an anti-hacking law, the CFAA is much broader in scope . . . it prohibits seven categories of conduct . . . .” Id. See also Justin Precht, The Computer Fraud and Abuse Act or the Modern Criminal at Work: The Dangers of Facebook from Your Cubicle, 82 CIN. L. REV. 359, at 360 (2014) (describing the 1986 amendment as remaining “limited to acts affecting federal government computers, but it expanded the CFAA to make it a criminal offense to ‘knowingly access a computer without authorization or to exceed authorized access.’”). See also 9-48.000 – COMPUTER FRAUD AND ABUSE ACT, U.S DOJ (Feb. 26, 2023), archived at https://perma.cc/LA46-39GM (summarizing the consultation policies and charging policies for CFAA cases).

\textsuperscript{15} See CFAA Background, supra note 12 (explaining the 1986 amendments). The 1986 amendment provided additional penalties for fraud and federal interest computers. Id. The CFAA also added three new provisions; section 1030(a)(4) prohibited unauthorized access with intent to defraud; section 1030(a)(5) prohibited accessing a computer without authorization and/or altering and destroying information; and section 1030(a)(6) prohibited trafficking computer passwords. Id. See also Berris, supra note 12 at 2 (listing seven categories of prohibited conduct under the CFAA). With certain exceptions and condition, the categories include:

(1) obtaining national security information through unauthorized computer access and sharing or retaining it; (2) obtaining certain types of information through unauthorized computer access; (3) trespassing in a government computer; (4) engaging in computer-based frauds through unauthorized computer access; (5) knowingly causing damage to certain computers by transmission of a program, information, code, or command; (6) trafficking in passwords or other means of unauthorized access to a computer; and (7) making extortionate threats to harm a computer or based on information obtained through unauthorized access to a computer.

Id.

\textsuperscript{16} See CFAA Background, supra note 12 (describing that CFAA had only provided criminal penalties until 1994). In 1994, “Congress added a civil cause of action for
the scope of conduct to prohibit transmissions of items that intentionally cause damage, expanding the application of the CFAA from a strict technical focus to include an actor’s malicious intent.\textsuperscript{17} In 1996, title II of the Economic Espionage Act broadened the application of the CFAA to all “protected computers,” which ultimately encompasses any computer connected to the internet.\textsuperscript{18} Post September 11, 2001, Congress enacted the Patriot Act, which expanded the definition of computers to include those physically outside the United States if they were used in connection with interstate or foreign commerce.\textsuperscript{19} Most recently, in 2008, Congress

CFAA violations that afforded private parties the ability to obtain compensatory damages, injunctive relief, and/or other equitable relief.” \textit{Id.}  
\textsuperscript{17} See \textit{id.} (explaining the 1994 amendments broadened the scope as related to transmissions). Section 1030(a)(5)(A) specifically prohibits “knowingly caus[ing] the transmission of a program, information, code, or command’ which ‘intentionally causes damage without authorization’ . . . the focus of the statute shifted from a technical concept of computer access and authorization, to the defendant’s malicious intent and resulting harm.” \textit{Id.}  
\textsuperscript{18} See Precht, \textit{supra} note 14, at 361 (explaining that “[t]he reach of the CFAA was expanded substantially by criminalizing the procurement of ‘information from any protected computer if the conduct involved an interstate or foreign communication,’ . . . by 1996 the CFAA protected a company’s proprietary information and could be used in actions by private employers against their employees.”); \textit{INTRODUCTION TO THE ECONOMIC ESPIONAGE ACT, THE U.S. DOJ ARCHIVES} (Oct. 22, 2022), archived at \url{https://perma.cc/P2PZ-ZYDU} (introducing the rationale for the economic espionage act).  

Congress, recognizing the importance of the protection of intellectual property and trade secrets to the economic health and security of the United States, enacted the Economic Espionage Act of 1996 . . . to address the growing problem of theft of trade secrets. The EEA contains two separate provisions that criminalize the theft or misappropriation of trade secrets. The first provision . . . is directed towards foreign economic espionage and requires that the theft of the trade secret be done to benefit a foreign government, instrumentality or agent. The second provision makes criminal the more common commercial theft of trade secrets, regardless of who benefits. 

\textit{Id.}  
\textsuperscript{19} See \textit{CFAA Background, supra} note 12 (stating that after the World Trade Center attacks in 2001, “[t]he most significant change was the expanded definition of ‘protected computer’ to include computers located \textit{outside} the United States; specifically, those computers ‘located outside the United States that [are] used in a manner that affects interstate or foreign commerce or communications of the United States.’”); \textit{USA PATRIOT Act, FIN. CRIMES EN’VT NETWORK} (Oct. 22, 2022), archived at \url{https://perma.cc/MHR8-BMJ7} (stating purpose of the Patriot Act “is to deter and punish terrorist acts in the United States and around the world, to enhance law enforcement investigatory tools, and other purposes. . . .”).
expanded the statute to its modern state, making it known as the all-encompassing “anti-hacking” law.20

B. A Brief Timeline of U.S. Data Privacy and Cyber Security Regulations

Unlike the European Union’s broad-reaching General Data Protection Regulation (“GDPR”), the United States’ data privacy regulations derive from a multitude of sources.21 Though it can be somewhat challenging for businesses to remain up-to-date and compliant with industry and state-specific laws, the Federal Trade Commission (“FTC”) provides some level of uniformity as the principal enforcer.22 Data privacy laws govern the collection, use, and

20 See CFAA Background, supra note 12 (indicating the recent amendments of the CFAA in 2008). Congress expanded the CFAA to “criminalize not only explicit threats to cause damage to a computer, but also threats to (1) steal data on a victim’s computer, (2) publicly disclose stolen data, or (3) not repair damage the offender already caused to the computer.” Id. Congress also created a criminal offense for conspiracy to commit a computer hacking offense under section 1030 and broadened the definition of “protected computer” by including those computers used in or affecting interstate or foreign commerce or communication. Id. See also Justices Clarify Scope of Anti-Hacking Law, JD SUPRA (June 10, 2021), archived at https://perma.cc/U3UV-TCWH (defining the CFAA “popularly known as the ‘anti-hacking’ law – makes it a crime for any person to knowingly access a computer without authorization or exceed the person’s authorized access to obtain information from a protected computer.”).

21 See David Harrington, U.S. Privacy Laws: The Complete Guide, VARONIS (Sept. 2, 2022) archived at https://perma.cc/U8Z5-YY5X (acknowledging “[t]he United States has a patchwork and ever-changing web of laws governing data privacy.”). GDPR came into effect in 2018 and “applies to any organization that processes or intends to process EU citizens’ sensitive data, regardless of the location . . . GDPR has a much broader reach and protection” than U.S. laws. Id. See also Michael Buckbee, What is the EU General Data Protection Regulation, VARONIS (Dec. 11, 2015), archived at https://perma.cc/YKG6-723K (describing that GDPR “legislates a lot of common sense data security ideas, especially from the Privacy by Design school of thought: minimize collection of personal data, delete personal data that’s no longer necessary, restrict access, and secure data through its entire lifecycle.”).

22 See Harrington, supra note 21 (articulating that “[t]he Federal Trade Commission (FTC) is the principal enforcer of these [data privacy] laws in the U.S. In recent years, the FTC has taken several enforcement actions against companies that have misled consumers about their data security and privacy practices.”). The FTC reached a settlement with Google, where they agreed to pay a $22.5 million fine and change their privacy practices. Id. In 2018, the FTC cracked down on Facebook for “deceiving users about their ability to control the visibility of their personal information.” Id. See also Federal Cybersecurity and Data Privacy Laws Directory, IT GOVERNANCE (Oct. 24, 2022), archived at https://perma.cc/Y9E4-2LVQ
disclosure of personal information within the U.S. and set standards for how businesses are required to handle this data.\textsuperscript{23}

1. Key Federal Privacy Focused Regulations

U.S. privacy laws generally fall into the categories of vertical or horizontal; while both are important in protecting privacy rights, many consider vertical as more effective in targeting specific types of risks.\textsuperscript{24} The earliest data privacy act passed by the federal government was the U.S. Privacy Act of 1974 ("Privacy Act"); it established rules and regulations surrounding the government’s use, collection, and disclosure of personal information.\textsuperscript{25} Noteworthy aspects of the Privacy Act include the right to request access and make corrections on any records maintained by the government that contain personal data about oneself.\textsuperscript{26}

(highlighting that “[s]everal states have their own cybersecurity and data breach notification laws” with respective summaries of each states’ guidelines). This can pose a “considerable challenge to organizations conducting business across all 50 states and worldwide.” Id.

\textsuperscript{23} See Harrington, supra note 21 (concluding that “data privacy laws govern the collection, use, and disclosure of personal data and set standards for how businesses need to handle sensitive data … the FTC is willing to crack down on companies that violate consumer privacy laws.”); Data Breach Notification Laws by State, IT GOVERNANCE (Oct. 24, 2022), archived at https://perma.cc/LR6K-UYJ5 (summarizing “[p]ersonal information in the United States is currently protected by a patchwork of industry-specific federal laws and state legislation whose scope and jurisdiction vary. The challenge of compliance for organizations that conduct business across all 50 states is therefore considerable.”).

\textsuperscript{24} See Harrington, supra note 21 (defining vertical privacy laws as those that “protect medical records or financial data, including details such as an individual’s health and financial status.”). Meanwhile, horizontal privacy laws “focus on how organizations use information, regardless of its context. The types of data covered by these laws include fingerprints, retina scans, biometric data, and other personally identifiable information such as names and addresses.” Id.

\textsuperscript{25} See The Privacy Act, HHS.GOV (Aug. 31, 2022), archived at https://perma.cc/564P-UHTE (outlining the Privacy Act of 1974 “[p]rotects records about individuals retrieved by personal identifiers such as a name, social security number, or other identifying number or symbol. An individual has rights under the Privacy Act to seek access to and request correction … or an accounting of disclosures of any such records” about him or her).

\textsuperscript{26} See Harrington, supra note 21 (summarizing the privacy act was intended to enhance individual privacy protection).

This act established rules and regulations regarding U.S. government agencies’ collection, use, and disclosure of personal information. … [E]xamples of the guaranteed rights covered by
Two highly sensitive areas, healthcare and finance, are governed by the Health Insurance Portability and Accountability Act (“HIPAA”) and the Gramm-Leach-Bliley Act (“GLBA”), respectively. HIPAA, enacted in 1996, is a federal law that created national standards to protect sensitive health information from being disclosed without the patient’s consent or knowledge. The privacy standard in HIPAA upholds safeguarding of protected health information (“PHI”) by covered entities; a major aspect of this is to ensure that individuals health information is properly secured while simultaneously permitting a certain level of information to flow in order to maintain the overall integrity of the healthcare system. On the financial side, the U.S. government signed the GLBA in 1999; it provides that any financial institution that collects, uses, or discloses personal information to explain their information sharing practices to individuals. The information privacy rule: [1] The right to request access and correct data if needed: U.S. citizens have a right to access their personal data kept by government agencies and request changes if they believe the information is inaccurate. [2] The right to access data (restricted on an individual basis): Government agencies grant users data access based on their role in their company. [3] The right to information about data uses: Individuals must know how agencies use their personal data upon collection.

Id. See Health Insurance Portability and Accountability Act of 1996 (HIPAA), CDC (Oct. 25, 2022), archived at https://perma.cc/Y6CF-FR9T (contending HIPAA was enacted to provide national healthcare standards for patient privacy); Harrington, supra note 21 (summarizing HIPAA and GLBA); Gramm-Leach-Bliley Act, FTC (Oct. 25, 2022), archived at https://perma.cc/XJ62-9YYM (covering “[t]he Gramm-Leach-Bliley Act requires financial institutions – companies that offer consumers financial products or services like loans, financial or investment advice, or insurance – to explain their information-sharing practices to their customers and to safeguard sensitive data.”).

27 See Health Insurance Portability and Accountability Act of 1996 (HIPAA), supra note 27 (providing that “[t]he Health Insurance and Accountability Act of 1996 (HIPPA) is a federal law that required the creation of national standards to protect sensitive patient health information from being disclosed without the patient’s consent or knowledge.”).

28 See id. (analyzing HIPAA Privacy Rule standards). The Privacy Rule addresses the use and disclosure of protected health information, also known as PHI. Id. It contains standards for individuals’ rights and allows them to understand how their health information is being used. Id. “A major goal of the Privacy Rule is to make sure that individuals’ health information is properly protected while allowing the flow of health information needed to provide and promote high-quality healthcare, and to protect the public’s health and well-being.” Id. See also Harrington, supra note 21 (listing rights under HIPAA). Covered entities can use patient data only for specific purposes, like treatment and payment. Id. Patients can also request restrictions on how providers disclose their information. Id.
their customers and safeguard their sensitive data. In essence, the law is both a privacy and information security law, and it applies to a broad range of financial institutions, including, banks, investment firms, securities firms, mortgage lenders, tax preparers, auto dealers, and insurance companies.

A third area of concern relates to minors, which both the Children’s Online Privacy Protection Act (“COPPA”) and the Family Educational Rights and Privacy Act (“FERPA”) govern in different variations. Enacted in 1988, COPPA applies to online services, websites, or mobile applications that collect information from individuals under the age of 13. Entities collecting such information, under COPPA, must publish a clear privacy policy about what information they collect and how they will use it; further, they must secure parental consent before securing personal data from children.

30 See Gramm-Leach-Bliley Act, 15 U.S.C. §§ 6801-6809, §§ 6821-6827 (providing codification of Gramm-Leach-Bliley Act). See also Gramm-Leach-Bliley Act, supra note 27 (providing overview of GLBA); Harrington, supra note 21 (highlighting that financial institutions must explain information-sharing practices to customers so they have the ability to opt out of data sharing, follow guidelines for how financial institutions can collect and use customer data, and develop a standardized information security program to protect customer data). See also Michael Buckbee, We Need to Talk About Gramm-Leach-Bliley (GLB): The Safeguards Rule Will Be Changing!, VARNIS (Mar. 29, 2020), archived at https://perma.cc/BT8D-28RP (detailing “[t]he GLBA’s data security regulations can be found in its Safeguards Rule, which in its current form provides very general requirements … ‘[y]ou shall develop, implement, and maintain a comprehensive information security program … contains administrative, technical, and physical safeguards that are appropriate to your size and complexity.’”).

31 See Federal Cybersecurity and Data Privacy Laws Directory, supra note 22 (referencing GLBA as “both an information security and a privacy law … [t]here is a Security Rule and a Privacy Rule.”). The Security Rule requires organizations to maintain an information security program appropriate to the business size, complexity, the nature of the scope of activities and the relative sensitivity of the consumer information being handled. Id.

32 See Harrington, supra note 21 (stating “[c]ongress enacted the Children’s Online Privacy Protection Act (COPPA) in 1998 to protect the online privacy of minors under the age of 13.”); Family Educational Rights and Privacy Act (FERPA), U.S. DEPT. OF EDUC. (Oct. 25, 2022), archived at https://perma.cc/LRD7-3DFA (explaining that FERPA is a “[f]ederal law that protects the privacy of student education records. The law applies to all schools that receive funds under an applicable program of the U.S. Department of Education.”).

33 See Student Privacy Laws: What District & School Administrators Need to Know, EDUC. FRAMEWORK (Oct. 25, 2022), archived at https://perma.cc/GW6U-FFHG (stating “COPPA applies to online services, commercial websites and mobile applications that, knowingly or unknowingly, collect information from individuals under 13.”).
and provide parents with access to review or delete their child’s information.\(^{34}\) Prior to the enactment of COPPA, FERPA was the primary law in protecting students education records since 1974.\(^{35}\) Under FERPA, parents and eligible students have the right to review the student’s education record, the right to amend the record, and the right for partial control over disclosure of the student’s personally identifiable information (“PII”) from their record.\(^{36}\)

\[^{34}\text{See id. (explaining that COPPA spells out what operators of sites and services must do for children’s online safety, which includes “impos[ing] certain requirements on those individuals that have access to student information.”). Id. “The Rule specifies what an operator needs to include in a privacy policy, what their responsibilities are to protect child privacy and safety online, and when they need to seek verifiable consent from a parent or guardian. The Rule also sets firm restrictions on marketing to children under 13.” Id. See also Harrington, supra note 21 (listing the steps under COPPA as: posting a clear and concise privacy policy explaining what information is collected and how it will be used, securing parental consent before using personal data from children, and providing parents with an opportunity to review their child’s personal information). See Josh Fruhlinger, COPPA explained: How this law protects children’s privacy, CSO (Feb. 8, 2021), archived at https://perma.cc/T69D-F3N3 (outlining when a site needs to comply with COPPA). You are regulated under COPPA if: (1) you run a website or service directed to children under 13 and collect their personal information, (2) you run ad networks or plug-ins that are collecting personal information from users under 13, (3) your website or service is directed to a general audience, but the provider has actual knowledge that personal information is being collected from children under the age of 13. Id. Whether a site is directed at children under 13 is assessed on a variety of criteria including, subject matter, visual and audio content, animation, presence of child celebrities or other evidence about the actual age of the intended audience. Id.}\]

\[^{35}\text{See Family Educational Rights and Privacy Act (FERPA), supra note 32 (summarizing FERPA and the level of rights parents and eligible students have with respect to the student’s education records).}\]

\[^{36}\text{See Student Privacy Laws: What District & School Administrators Need to Know, supra note 33 (stating the general level of involvement parents have over their child’s education record). FERPA allows parents to access their child’s education record, to amend the record, and to have control over disclosure of PII from the record. Id. Schools are expected to respond within 45 days of a parent requesting information related to their child’s record. Id. See also Chitra Mittha, Student data privacy in schools, ADOBE BLOG (Oct. 4, 2021), archived at https://perma.cc/38KQ-YRCD (describing personally identifiable information (PII) as “includ[ing] all info that could be used to identify, locate or contact that student – such as their full name, address or social security number.”). “Schools and digital platforms are also restricted from sharing students’ academic, health, and disciplinary records with any third party who’s not authorized to access them.” Id.}\]
2. Key Federal Cyber Focused Regulations

The United States’ federal role in addressing cybersecurity is similarly complex and even intertwined with federal regulations that address data privacy. There is no overarching regulation that guides the federal and nonfederal systems; instead, there is an array of enacted statutes that each address different aspects of cybersecurity.

The Electronic Communications Privacy Act (“ECPA”) was enacted in 1986 to update the Federal Wiretap Act of 1968 and ensure that legislation was evolving alongside the progression of communication methods and technologies. Generally, the ECPA protects wire, oral, and electronic communications that are being made, in transit, or digitally stored. The ECPA is separated into three specific provisions to address the different types of communication, but ultimately applies to email, telephone, and electronically stored data.

---


38 See id. (highlighting there is no overarching framework legislation for cybersecurity in place, but rather many enacted statutes that each address different areas of cyber).


40 See Electronic Communications Privacy Act of 1986 (ECPA), supra note 39 (clarifying the ECPA “amended, protects wire, oral, and electronic communications while those communications are being made, are in transit, and when they are stored on computers.”).

41 See id. (listing three titles of ECPA and referencing “[t]he act applies to email, telephone conversations, and data stored electronically.”). Title I of the ECPA is the Wiretap Act and Title II is the Stored Communications Act, which “protects the privacy of the contents of files stored by service providers and of records held about the subscriber by service providers, such as subscriber name, billing records, or IP addresses.” Id. Title III “addresses pen register and trap and trace devices . . . (a device that captures the dialed numbers and related information to which outgoing calls or communications are made by the subject[])” Id. See also Electronic Communications Privacy Act (ECPA), Elec. Priv. Info. Ctr. (Oct. 26, 2022),
In 2002, the Department of Homeland Security ("DHS") was given discretion over certain cybersecurity responsibilities through the Homeland Security Act ("HSA"). One of the provisions within the HSA is the Cyber Security Enhancement Act (CSEA), which loosened restrictions on internet service providers and the threshold of their legal authority to divulge communications and data pertaining to national security interests. In the same year, the Sarbanes-Oxley Act ("SOX") was enacted to improve the transparency and reliability of disclosures and financial statements for all U.S. publicly traded companies. Though SOX’s focus derived from financial reporting, there are two key relevant cybersecurity provisions directed towards companies internal controls. These provisions require that...
companies’ infrastructure and applications protect financial data, ensure the financial data is correct, and report any breaches or incidents in a timely manner.46

3. Variations of Privacy Laws at State Level

Given the lack of a single federal data privacy law in the United States, companies have ample opportunity to use consumer data how they see fit.47 If the data they are utilizing falls outside of the select areas of health, finance, or minors, it could be entirely unregulated based on the state(s) they are operating in.48 The majority of states permit companies to sell, use, and share consumer data without notification; if this data is breached, the consumer may not even know.49

In 2018, the California Consumer Privacy Act (“CCPA”) was signed into law as the first major consumer privacy law in the United States. Section 302 requires “companies to have internal controls which ensure accurate, timely financial disclosures. The controls have to make sure data is accurate and remains secret until it’s disclosed.” Id. “Section 404, which requires internal controls to protect financial data and ensure proper reporting.” Id. 46 See id. (generalizing that “SOX cybersecurity compliance is about companies implementing robust, internal controls to protect financial information and related financial reports in company infrastructure and applications.”). 47 See Thorin Klosowski, The State of Consumer Data Privacy Laws in the US (And Why It Matters), WIRECUTTER (Sept. 6, 2021), archived at https://perma.cc/6WMD-BBWB (describing privacy laws currently as a “cluttered mess of different sectoral rules.”). Amie Stepanovich, executive director at the Silicon Flatirons Center at Colorado Law states, “[h]istorically, in the US we have a bunch of disparate federal [and state] laws . . . .” Id. 48 See id. (describing the specific laws in the US as “either look[ing] at specific types of data, like credit data or health information . . . or look[ing] at specific populations like children, and regulat[ing] within those realms.”). “The United States . . . doesn’t have a singular law that covers the privacy of all types of data. Instead, it has a mix of laws that go by acronyms like HIPAA, FCRA, FERPA, GLBA, ECPA, COPPA, and VPPA . . . .” Id. Majority of the data collected doesn’t necessarily fall into these stringent categories, thus it is not regulated. Id. 49 See Klosowski, supra note 47 (stating companies are basically free to do what they want with consumers’ data).

In most states, companies can use, share, or sell any data they collect about you without notifying you that they’re doing so. No national law standardizes when (or if) a company must notify you if your data is breached or exposed to unauthorized parties. If a company shares your data, including sensitive information such as your health or location, with third parties (like data brokers), those third parties can further sell it or share it without notifying you.

Id.
The CCPA intends to protect consumers by giving them the right to request what type of personal information is collected, the source of information, and the business purpose for collecting their information.\(^{51}\) In 2020, the CCPA was expanded to include the California Consumer Privacy Rights Act (“CPRA”), which establishes the right for consumers to prevent, limit, or correct how businesses are sharing their personal information.\(^{52}\) Both the CCPA and CPRA apply to organizations that “do business” in California and meet certain thresholds related to annual gross revenue, number of consumers, and level of revenue derived from consumer data.\(^{53}\) In the event of non-compliance, organizations that do business in California could face fines ranging from $2,500 for unintentional violations up to $7,500 for intentional violations.\(^{54}\) If the consumer brings an action directly, the

\(^{50}\) See US State Privacy Legislation Tracker, IAPP (Oct. 7, 2022), archived at https://perma.cc/U46Y-YHUZ (identifying the CCPA was signed into law in 2018, to become effective on January 1, 2020). See also Tony Romm, California adopted the country’s first major consumer privacy law. Now, Silicon Valley is trying to rewrite it., WASH. POST (Sept. 3, 2019), archived at https://perma.cc/68JN-MVPY (stating California was the first state to adopt a major consumer privacy law).

\(^{51}\) See State Laws Related to Digital Privacy, NCSL (June 7, 2022), archived at https://perma.cc/J8YT-347E (stating the CCPA of 2018 “[a]llows consumers the right to request a business to disclose the categories and specific pieces of personal information that the business has collected about the consumers as well as the source of that information and business purpose for collecting the information.”). See also Data Privacy Laws by State: Comparison Charts, BLOOMBERG L. (Feb. 2, 2022), archived at https://perma.cc/2HCX-NB3A [hereinafter Data Privacy Laws by State] (defining personal information as “[i]nformation 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.”) Consumer is defined as a “natural person who is a California resident.” Id. Consumer is defined as a “natural person who is a California resident.”

\(^{52}\) See State Laws Related to Digital Privacy, supra note 51 (stating CPRA expands the consumer data privacy laws within CCPA). CPRA “[p]ermits consumers to: (1) prevent businesses from sharing personal information; (2) correct inaccurate personal information; and (3) limit businesses’ use of ‘sensitive personal information’—including precise geolocation; race; ethnicity; religion; genetic data; private communications; sexual orientation; and specified health information.” Id. The CPRA was approved November 2020 and becomes effective law in January 2023. Id.

\(^{53}\) See Data Privacy Laws by State, supra note 51 (explaining that the CCPA and CPRA apply to “entities that ‘do business’ in California that meet the following thresholds: [1] annual gross revenues greater than $25 million, [2] process the data of 50,000 or more consumers, [3] at least 50% of revenue comes from selling of data.”).

\(^{54}\) See id. (indicating that for both the CCPA and CPRA the consequences of non-compliance are fines “up to $7,500 per intentional violation or $2,500 per unintentional violation.”).
organization could be penalized between $100 and $750 per incident, or based on the direct amount of damage.\textsuperscript{55}

Though California is leading the way with the strongest privacy laws in the United States, Colorado, Virginia, Connecticut, and Utah have committed to enacting their own legislation.\textsuperscript{56} Colorado’s Privacy Act (“CPA”) and Virginia’s Consumer Data Protection Act (“VCDPA”) are similar to the CCPA in that they protect consumers who are residents of their respective states and apply to businesses that meet certain thresholds; however, these acts apply to organizations that “conduct business” in the state or produce products and services that “target” their residents.\textsuperscript{57} Utah’s Consumer Privacy

\textsuperscript{55} See id. (explaining that for both the CCPA and CPRA, “[i]n actions brought by consumers for security breach violations, statutory damages not less than $100 and not greater than $750 per consumer per incident or actual damages, whichever is greater.”).

\textsuperscript{56} See Klosowski, supra note 47 (declaring that some experts consider California’s privacy protections as the strongest in the U.S.). California’s privacy regulations include a limited private right of action, which is the ability to sue a company against a data breach as an individual consumer. Id. Additionally, California requires that a “global opt out” so that consumers do not have to opt out of data sharing on each individual site. Id. Virginia and Colorado are among the states that have comprehensive consumer privacy laws, with Massachusetts, New York, North Carolina, and Pennsylvania as having proposals in committee. Id. There is a risk that with too many state laws, there is great confusion for consumers and operationally for companies. Id. See also State Laws Related to Digital Privacy, supra note 51 (describing overview of Colorado, Connecticut, Utah, and Virginia consumer privacy laws). See also US State Privacy Legislation Tracker, supra note 50 (demonstrating that Colorado’s Privacy Act was signed in 2021 to become effective in July 2023, Connecticu’t was signed in 2022 to become effective in July 2023, Virginia’s was signed in 2021 to become effective in January 2023, and Utah’s was signed in 2022 to become effective in December 2023).

\textsuperscript{57} See Data Privacy Laws by State, supra note 51 (summarizing that the VCDPA protects consumers, defined as a “[n]atural person who is a VA resident” and the CPA protects consumers, defined as an “[i]ndividual who is a CO resident[]”). The type of data protected by both of these regulations is personal data, “information that is linked or reasonably linkable to an identified or identifiable natural person.” Id. “VCDPA applies to entities that ‘conduct business’ in Virginia or produce products or services ‘targeted’ to Virginia residents.” Id. The VCDPA has no revenue threshold; it only applies to entities that process the data of more than 100,000 consumers or companies that process the data of a minimum of 25,000 consumers when they retrieve fifty percent of their revenue from the sale of such data. Id. The “CPA applies to any entity that ‘conducts business’ in Colorado or produces or delivers commercial products or services ‘intentionally targeted’ to Colorado residents.” Id. One of two thresholds exist in the Colorado Privacy Act; organizations must either control or process the data of a minimum of 100,000 consumers or derive revenue from the personal data of at least 25,000 consumers. Id.
Act ("UCPA") is quite similar in that it provides consumers transparency on how their data is being used and whether a business sells their data.\textsuperscript{58} Additionally, the UCPA applies to for-profit organizations that have an annual revenue of $25 million and that meet one of two thresholds related to (1) the number of consumers or (2) the amount of gross revenue from selling personal data.\textsuperscript{59} Finally, Connecticut’s Personal Data Privacy and Online Monitoring act establishes standards for controlling and processing data and instills that consumers have the right to access, delete, correct, and opt out of personal data processing; this act applies to any organization that collects personal information from Connecticut residents.\textsuperscript{60}

4. The General Data Protection Regulation

In 2018, the European Union ("EU") enacted the General Data Protection Regulation ("GDPR") into law.\textsuperscript{61} GDPR is the most

\textsuperscript{58} See State Laws Related to Digital Privacy, supra note 51 (stating the UCPA provides consumers “the right to know what personal data a business collects, how the business uses the personal data, and whether the business sells the personal data.”). “It also requires specified businesses to safeguard personal data, provide clear information about how consumers’ personal data [is] used, and accept and comply with consumer requests to access, delete or stop selling personal data.” Id.

\textsuperscript{59} See Kate Berry et al., And Utah Makes 4–Beehive State Passes Consumer Privacy Law, DAVIS WRIGHT TREMAINE LLP (Mar. 31, 2022), archived at https://perma.cc/25ES-77S8 (outlining scope of the UCPA).

The UCPA applies to for-profit entities (“controllers” or “processors” that (1) conduct business in Utah or target products and services to consumers who are residents of the state, (2) have annual revenues of at least $25 million, and (3) meet one of two threshold requirements: [1] [a]nnually control or process the personal data of 100,000 or more Utah residents (“consumers”); or [2] [d]erive over 50 percent of gross revenue from the “sale” of personal data and control or process personal data of 25,000 or more consumers.

Id.

\textsuperscript{60} See Harrington, supra note 21 (describing Connecticut’s act as “cover[ing] any business that collects personal information from Connecticut residents.”). See also State Laws Related to Digital Privacy, supra note 51 (addressing “[t]he Connecticut act establishes a framework for controlling and processing personal data; provides responsibilities and privacy protection standards for data controllers and processors; and grants consumers the right to access, correct, delete and obtain a copy of personal data, and opt out of the processing of personal data.”).

\textsuperscript{61} See What is GDPR, the EU’s new data protection law?, GDPR.EU (Nov. 20, 2022), archived at https://perma.cc/7VS-K-NSLM (distinguishing “[t]he GDPR entered into force in 2016 after passing European Parliament, and as of May 25, 2018, all organizations were required to be compliant.”).
thorough and strict privacy and security law for organizations to comply with; organizations worldwide are obliged to comply so long as they target or collect personal data related to EU residents, regardless of whether they operate within the EU. The GDPR is detailed; it outlines data protection principles, addresses appropriate measures for data security, lists guidelines for legally processing personal data, and requires controllers to demonstrate compliance. For example, the GDPR even outlines specific conditions in which organizations should have a Data Protection Officer appointed. Furthermore, its penalties for non-compliance are severe; companies that violate the regulation are subject to “administrative fines up to 20 million or 4% of total worldwide annual [revenue] of the preceding financial year, whichever is higher.”

See id. (describing GDPR as “the toughest privacy and security law in the world.”). “Though it was drafted and passed by the European Union (EU), it imposes obligations onto organizations anywhere, so long as they target or collect data related to people in the EU.” Id. Europe is utilizing the GDPR to signal its strong stance on data privacy at a critical time when individuals are entrusting organizations daily with handling and storing their personal data. Id. See also Data Privacy Laws by State, supra note 51 (acknowledging GDPR requires compliance by “any entity that processes personal data in the context of activities of an ‘establishment’ in the EU, or processes personal data of individuals in the EU related to the offering of goods and services to them or monitoring their behavior.”). There is no revenue threshold or processing amount threshold that organizations have to meet to be subject to GDPR; it is very broad reaching. Id.

See What is GDPR, the EU’s new data protection law?, supra note 61 (summarizing the key regulatory points of GDPR). The GDPR outlines that in order to process data, you have to follow seven protection and accountability principles. Id. “[D]ata controllers have to be able to demonstrate they are GDPR compliant.” Id. To handle data securely, organizations must implement appropriate technical measures, which means “anything from requiring your employees to use two-factor authentication on accounts where personal data are stored to contracting with cloud providers that use end-to-end encryption.” Id. See also Recital 78, GDPR.EU (Feb. 26, 2023), archived at https://perma.cc/EG85-TRM5 (providing the appropriate technical and organizational measures required by GDPR).

See What is GDPR, the EU’s new data protection law?, supra note 61 (defining conditions when an organization needs a data protection officer). The conditions include (1) being a public authority acting in a judicial capacity, (2) if the core activities of the business require the company to monitor people regularly on a large scale, (3) the core activities of large-scale processing falls within Article 9 or 10 of the GDPR. Id.

See id. (describing the fines for violating GDPR as extremely high). Two tiers of penalties exist: companies could be subject to a fine of 20 million euros or four percent of global revenue, whichever is higher. Id. In addition to the standard fines imposed by the regulation, individuals also have the right to seek compensation for damages. Id.
global companies have been found non-compliant with various GDPR provisions, resulting in hundreds of millions of dollars in fines.66

C. Historical Use of Big Data & Emergence of Web Scraping

People have been using data analysis for many centuries, but the term “Big Data” emerged in the early 1990s.67 It is the all-encompassing term for large sets of information generated by people, machines, and tools.68 Big Data originated from the domain of database management, which was common in the late 1990s and consisted of dashboards, scorecards, data mining, and statistical analysis.69 Big Data phase 2.0 arrived in the early 2000s when websites began to analyze customer behavior; this disrupted the previous storage solutions and forced companies to restructure their

See also Data Privacy Laws by State, supra note 51 (describing the consequences of non-compliance with GDPR).
66 See 30 Biggest GDPR Fines So Far (2020, 2021, 2022), TESSIAN (May 5, 2022), archived at https://perma.cc/3YDX-PT92 (stating since the GDPR took effect in 2018, over 900 fines have been issued). In July 2021, Amazon was fined $877 million for reasons related to their cookie consent. Id. In 2022, Facebook received a $68 million fine because their cookie consent for users was deemed to be unclear. Id.
67 See Where does ‘Big Data’ come from?, ENTER. BIG DATA FRAMEWORK (Mar. 26, 2019), archived at https://perma.cc/SG4G-Y5NS (explaining “[t]he term ‘Big Data’ has been in use since the early 1990s. . . . Big Data is not something that is completely new or only of the last two decades.”). “Over the course of centuries, people have been trying to use data analysis and analytics techniques to support their decision-making process.” Id. “The ancient Egyptians around 300 BC already tried to capture all existing ‘data’ in the library of Alexandria. Moreover, the Roman Empire used to carefully analyze statistics of their military to determine the optimal distribution for their armies.” Id.
68 See William Goddard, What is Big Data And How Does it Work?, IT CHRON.’S (Mar. 23, 2021), archived at https://perma.cc/88LX-8HKZ (defining big data as “a blanket term for the dynamic, often extremely large sets of information generated by people, machines, and tools. Big data sources encompass information from social media, machine data, smartphones, tablets, video, voice recordings, and the preservation and logging of structured and unstructured data.”).
69 See Where does ‘Big Data’ come from?, supra note 67 (stating “[d]ata analysis, data analytics and Big Data originate from the longstanding domain of database management.”). Big Data “relies heavily on the storage, extraction, and optimization techniques that are common in data that is stored in Relational Database Management Systems (RDBMS).” Id.
data storage mechanisms. The core characteristics associated with Big Data were volume, variety, and velocity; however, this began to shift in 2010 with the emergence of the mobile era, known as Big Data 3.0.

Around the same time that Big Data phase 1.0 took off, the concept of web scraping emerged. Shortly after the creation of the World Wide Web, the first web crawler, the Wanderer, was created to measure the size of the web. In 1993, JumpStation, a crawler-based web search engine was created; it indexed webpages using keywords and displayed results as URLs, but did not extract any of the data. By the early 2000s, BeautifulSoup emerged and allowed programmers to extract data from HTML and XML files. With the expansion of

---

70 See id. (depicting a table that shows Big Data Phase 2.0 lasted from 2000-2010). “Since the early 2000s, the Internet and the Web began to offer unique data collections and data analysis opportunities . . . From a data analysis . . . point of view, HTTP-based web traffic introduced a massive increase in semi-structured and unstructured data.”). Id.

71 See Goddard, supra note 68 (highlighting that “[c]onceptually, big data has certain core characteristics, most commonly denoted by the three Vs[.]”). “[1] Volume: [t]he vast amounts of information created in comparison to traditional data sources. [2] Variety: [b]ig data comes from a multiplicity of sources and can take numerous formats. [3] Velocity: [d]ata is generated at rapid speeds and must be handled at a similarly fast pace.” Id. Recently, analysts are beginning to add veracity and value to the list. Id. See also Where does ‘Big Data’ come from?, supra note 67 (depicting that big data phase 3.0 started in 2010 and remains current today).

72 See Web Scraping History: The Origins of Web Scraping, SCRAPINGROBOT (Apr. 8, 2022), archived at https://perma.cc/K2K7-7H5H (stating web scraping can be traced back to the beginning of the internet); Where does ‘Big Data’ come from?, supra note 67 (displaying big data phase 1.0 can be traced back as early as 1970 and lasted until 2000).

73 See Web Scraping History: The Origins of Web Scraping, supra note 72 (discussing that after the world wide web was created, the “World Wide Web Wanderer, was born in 1993. Created by Matthew Gray at [MIT], the Wanderer was a Perl-based web crawler that measured the size of the World Wide Web.”).

74 See id. (stating that JumpStation was born in 1993). JumpStation “indexed millions of web pages, turning the internet into an expansive open-source platform the world had never seen before.” Id. “It used headings and document titles to index web pages found through a simple linear search . . . also showed its results in the form of URLs that matched the users’ keywords. . . . [but] wasn’t intended to pull massive amounts of data from websites.” Id.

75 See id. (stating BeautifulSoup emerged in 2004). See also Jeri Wieringa, Intro to Beautiful Soup, PROGRAMMING HISTORIAN (Dec. 30, 2012), archived at https://perma.cc/S4HE-GR8S (describing “Beautiful Soup is a Python library for getting data out of HTML, XML, and other markup languages. . . . [It] helps you pull particular content from a webpage, remove the HTML markup, and save the information.”)
internet access, this became very common. However, because it was still quite difficult for non-programmers to utilize these data extraction methodologies, modern web-scraping was born.

III. Facts

A. Modern Use of Big Data

Data is generated at a rapid rate from millions of different sources and in a variety of formats including: pictures, excel spreadsheets, emails, google searches, phone calls, and social media feeds. Harnessing the power of big data is a critical component of maintaining a competitive edge in the modern business landscape; this requires the adoption of machine learning algorithms and big data analysis to truly understand what the data means and how the data’s value can serve the business. When companies process and understand these data sets, their innovation multiplies; they are able to predict market trends, find patterns and correlations related to

---

76 Web Scraping History: The Origins of Web Scraping, supra note 72 (describing that by the time the internet became readily accessible, “many people started using BeautifulSoup to extract text, pictures, and other information from the web.”).

77 See id. (explaining that one has to know how to code or program to use BeautifulSoup).

78 See Louie Andre, 53 Important Statistics About How Much Data Is Created Every Day, FINANCESONLINE (Nov. 4, 2022), archived at https://perma.cc/W9GZ-HMMH (displaying the amount of data generated by the minute, data growth in the year of 2021, and estimated data consumption from 2021 to 2024); Simplilearn, What is Big Data Analytics and Why It is Important?, SIMPLILEARN SOL.’s (Sept. 28, 2022), archived at https://perma.cc/8MKQ-TXZ8 (describing that big data contains massive amounts of data which cannot be stored, processed or analyzed using traditional data techniques). “Today, there are millions of data sources that generate data at a very rapid rate.” Id. See also Eilis McCann, Understanding the Power of Big Data in Business, METHOD INTEGRATION (Nov. 9, 2022), archived at https://perma.cc/X4CB-DUFJ (listing potential sources of data include: website clickstream behavior, mobile apps, social media accounts, and accounting or other internal software).

79 See Nikita Duggal, Top 7 Benefits of Big Data and Analytics and Reasons to Consider It For Your Next Career Move, SIMPLILEARN SOL.’s (Oct. 17, 2022), archived at https://perma.cc/43SM-6MMH (describing big data as an essential requirement for businesses to harness their true potential). Big businesses like Amazon, American Express, Netflix, Capital One, and more, are using the power of big data analytics to gain success. Id. See also Simplilearn, supra note 78 (setting forth “[t]oday, Big Data analytics has become an essential tool for organizations of all sizes across a wide range of industries.”). See also Glikman, supra note 4 (highlighting “[d]ata has become a strategic asset that allows companies to acquire or maintain a competitive edge.”).
consumer preferences, optimize efficiency and costs, and enhance complex systems. For example, Amazon has optimized big data analytics to improve customers’ experiences; they collect information on whether the customer is simply looking at items, adding items to their cart, or ultimately purchasing them to improve product suggestions and update their pricing strategy accordingly. Additionally, Netflix has been able to harness a 93% user retention rate (the highest in the industry) by optimizing big data.

A common theme within the power of big data is personalization; companies are highly interested in learning more about their individual users to better enhance their experience and ultimately make a profit. Some of the most prevalent firms in the world are greatly valued due to the consumer data they own.

80 See Duggal, supra note 79 (listing the most compelling benefits of big data as: customer acquisition and retention, targeted promotions, risk identification, innovation, cost optimization, navigating supplier networks, and improved efficiency). See also Simplilearn, supra note 78 (evaluating “[b]ig data analytics . . . reveal[s] information – such as hidden patterns, correlations, market trends, and consumer preferences – that may assist businesses in making educated business choices.”).

81 See Nina Tudor, 7 real-world examples how brands are using Big Data analytics, BORNFIGHT (Apr. 27, 2020), archived at https://perma.cc/8X47-6XPE (recognizing Amazon as the number one e-commerce shop constantly using big data to improve their customer).

It doesn’t matter if the person buys the products, puts it in the cart or even just takes a look at it – Amazon will use that data … they learn what each customer wants and likes, and can recommend that same product or similar ones to them when they return to the shop. Id. Ultimately, this use of big data accounts for roughly thirty-five percent of Amazon’s annual sales. Id.

82 See Tudor, supra note 81 (stating “Netflix is unarguably the biggest online platform for streaming movies and TV shows, and it owes its success to Big Data . . . their retention rate is 93% which, compared to their main competitors, is a huge number.”). Netflix collects data on when users subscribe to a show, whether they binge-watched it or took time to finish the series, and even if they paused the show and continued to watch it.

83 See Freedman, supra note 5 (describing “[m]apping users’ journeys and personalizing their journey . . . is now essential.”). The demand for data is increasing; “[c]ompanies that capture data stand to profit from it.” Id.

84 See Betsy Vereckey, New research puts a price on the value of financial data, MIT (Feb. 23, 2022), archived at https://perma.cc/CLK8-D84Z (claiming “[r]ight now, some of the largest firms in the U.S. are valued heavily for the data they own, such as consumer data or production data.”). Companies purchase data to make smarter decisions, “but only if the value they get from that data is greater than the price they paid for it.” Id. “Companies sell data, too – but they need to understand the value of the data so that they can put a price on it.” Id. See also Tudor, supra note 81
valuation can take on a number of forms: valuation by the shareholder, by the function of the companies’ user base, and by the individual user themselves.\textsuperscript{85} Although the value of data fluctuates based on the unique situation of who and how it is being used, the value of raw consumer data often falls below $1.00.\textsuperscript{86} Companies are purchasing and monetizing this data for a financial return, whether that occurs directly, by selling the data to a third party, or internally through improved target marketing campaigns that boost sales.\textsuperscript{87} Consumer data is a top priority for companies across an array of industries, but this data belongs to individual people, who are often times left in the dark on how it is being used.\textsuperscript{88}

(providing descriptions of how Amazon, Marriott, McDonalds, Starbucks, Uber, and Netflix are leveraging consumer data).

\textsuperscript{85} See Glikman, \textit{supra} note 4 (listing different forms of data valuation).

\textsuperscript{86} See id. (stating there are currently 270 data brokers worldwide and the value of the data sold is often below $1). “For example, general information about an individual (age, sex, locality) costs only $0.0007.” \textit{Id}. “For the shareholder, data embodies a financial potential. For the company, data can be used to optimize the way it does business . . . the value of personal data still depends largely on who uses it, how it is being used and in what context.” \textit{Id}.

\textsuperscript{87} See Jeffrey Brashear, et al., \textit{Data as a strategic asset}, \textit{DELOITTE} (Nov. 17, 2022), archived at https://perma.cc/Y6D3-36MK (finding “[a] data-driven enterprise can be positioned to monetize data in multiple ways.”). This can include “direct revenue in the form of data assets sold to relevant third-parties . . . .” \textit{Id}. The bigger value can lie in the form of “data packages and combinations that advance the business . . . .” \textit{Id}.

\textsuperscript{88} See Andrew Busby, \textit{New PwC Survey Reveals Consumer Data Is The Most Highly Valued}, \textit{FORBES} (Mar. 4, 2019), archived at https://perma.cc/ZE5T-DXWB (claiming consumer data is the data most businesses want). PwC survey indicates that consumer data is at the top of the list for companies to extract value from. \textit{Id}. See also Glikman, \textit{supra} note 4 (acknowledging “despite large-scale outrage over data breaches, it looks as if the only part of the data chain that is not willing to pay in order to protect it is paradoxically the individual himself – who happens to be its source.”). See also Morey, \textit{supra} note 4 (showing percentages of people that are in the dark about data). Although most people are aware that companies collect information about them, they are uninformed about the specific types of data they give up when browsing the internet. \textit{Id}. About twenty-five percent of people realize their location is being shared, while only fourteen percent their web-surfing history is being shared. \textit{Id}.
B. Data Scraping Under the CFAA

1. Evolving Interpretations of the CFAA

Today, the CFAA recognizes two forms of liability related to unauthorized access to a computer: “without authorization” and “exceed[ing] authorized access.”89 The statute is limited in terms of defining these phrases; there is great ambiguity regarding what it means to “access” a computer and what it means to do so “without authorization.”90 Though “exceeds authorized access” is defined in the statute as “[t]o access a computer without authorization and to use such access to obtain or alter information in the computer that the accesser is not entitled so to obtain or alter,” courts have struggled to interpret and apply this language.91

89 See 18 U.S.C. § 1030(a)(1) (2012) (codifying “having knowingly accessed a computer without authorization or exceeding authorized access . . . .”). See also § 1030(a)(2) (stating “intentionally accesses a computer without authorization or exceeds authorized access”). See also § 1030(a)(4) (stating “knowingly and with intent to defraud, accesses a protected computer without authorization, or exceeds authorized access . . . .”). See also Patricia L. Bellia, A Code-Based Approach to Unauthorized Access Under the Computer Fraud Abuse Act, 84 GEO. WASH. L. REV. 1142, 1444 (2016) (pointing that the “CFAA currently encompasses two forms of unauthorized access to a computer: ‘access[ing]’ a computer ‘without authorization’ and ‘exceed[ing]’ authorized access.”)

90 See Bellia, supra note 89 (explaining “[t]he statute does not define what it means to ‘access’ a computer or what it means to do so ‘without authorization.’” See also Aaron Mackey & Kurt Opshal, Van Buren is a Victory Against Overbroad Interpretations of the CFAA, and Protects Security Researchers, ELEC. FRONTIER FOUND. (June 3, 2021), archived at https://perma.cc/C6TV-A88P (referring to the CFAA definition and stating that it “does not define what authorization means for purposes of exceeding authorized access.”). A gates-up-or-down approach means that “either you are entitled to access the information or you are not. If you need to break through a digital gate to get in, entry is a crime, but if you are allowed through an open gateway, it’s not a crime to be inside.” Id.

91 See 18 U.S.C. § 1030(6) (2012) (defining term “exceeds authorized access” as “to access a computer with authorization and to use such access to obtain or alter information in the computer that the accesser is not entitled so to obtain or alter . . . .”). See also Bellia, supra note 89 (defining exceeds authorized access within the CFAA and stating that “[c]ourts have long struggled to apply these concepts of accessing a computer without authorization and exceeding authorized access . . . [i]ndeed, the case law reflects at least five different interpretative paradigms.”). See also Ambrose V. McCall, Which Rule of Statutory Interpretation Applies to the Computer Fraud and Abuse Act, THE FED. LAW. (Aug. 2011), archived at https://perma.cc/W7K8-Y63L (describing the variations in applications of appellate court opinions).
The circuit courts were historically split on how to interpret “authorization,” with the Fourth and Ninth Circuit courts adopting a narrow interpretation of authorization. The court focused on the technical methods of obtaining data; “without authorization” was applied to outside hackers who had no authorized access to the computer and “exceeds authorized access” was applied to inside hackers “whose initial access to a computer is authorized but who access[es] unauthorized information or files.” Often times, this is

The appellate courts have not sung in unison or harmony when deciding what constitutes unauthorized access of computers or even which test to apply under the CFAA. For example, the Seventh Circuit strongly contends that common law agency principles apply when determining whether a person is authorized to access a protected computer. In contrast, the Ninth Circuit dismisses the use of common law agency principles and relies on a plain reading of the CFAA. . . . Until resolved by Congress or the Supreme Court, counsel must determine which interpretation applies in the jurisdiction in which they practice while keeping in mind the plain meaning and intended use analyses.

Id.

92 See Din, supra note 12, at 418–19 (alluding to differences in the courts application and interpretation of “authorized.”). “[P]robably, the word ‘authorized’ is never defined. Consequently, circuits have interpreted the term ‘authorization’ differently depending on the context of the case; as a result, the statute has not been consistently interpreted … The Fourth and Ninth Circuits have adopted a narrow definition of authorization.” Id. See also Precht, supra note 14, at 364–65 (explaining that courts, including the Ninth Circuit, “have rejected the broad interpretation and adhere to a narrow interpretation of the CFAA that refuses to criminalize violations of employer computer use policies.”). “[T]he act was originally designed to target hackers who accessed computers to steal information or to disrupt or destroy computer functionality, as well as criminals who possessed the capacity to ‘access and control high technology processes vital to our everyday lives.’” Id.

93 See Berris, supra note 12, at 7 (stating that “[e]ven if the meanings of ‘exceeds authorized access’ and ‘without authorization’ are unclear, there is some indication in legislative history that the two phrases were intended to correspond to different categories of unauthorized computer use.”). “Without authorization” was originally intended to apply to outside hackers who had no authority to access the computer in the first place; while, “exceeds authorized access” was probably meant to apply to insider employees who surpassed their permissions. Id. See also Din, supra note 12, at 419 (explaining the narrow interpretation applied by the Fourth and Ninth Circuit courts). The courts interpreted the “phrase ‘without authorization’ as designed to apply to outside hackers who have no authorized access to a computer and the phrase ‘exceeds authorized access’ to apply to inside hackers ‘whose initial access to a computer is authorized but who access unauthorized information or files.’” Id. The court focused on the “technical means by which data was obtained. Insiders would be hackers only if they obtained data to which they did not have
subject to whether a code-based barrier is involved. The courts rest
their narrow interpretation on the rule of lenity, which provides that
for ambiguous criminal statutes, the courts should rule the ambiguity
in favor of the defendant to defer penalizing those whose conduct does
not create risks of harm that the statute was originally intended for.
Ultimately, the court expressed concerns that the broader application
of the statute would criminalize large groups of people who were not
actually committing a crime.

precise access, even if they had access to the broader network where such data was
stored.” Id. See also WEC Carolina Energy Solutions LLC v. Miller, 687 F.3d 199,
206 (2012) (holding the Fourth Circuit adopts a narrow reading of both “without
authorization” and “exceeds authorized access.”). “[T]hey apply only when an
individual accesses a computer without permission or obtains or alters information
on a computer beyond that which he is authorized to access.” Id.

94 See Bellia, supra note 89, at 1457 (referencing that “[u]nder the code paradigm,
whether access to a computer is unauthorized depends upon whether an individual
breaches a code-based barrier to the system or to certain information on it.”).
Typically, code-based applications involve “situations in which an employee has
technical access to a computer system and the employer has no policy that speaks
clearly to the question of access.” Id.

95 See Din, supra note 12, at 420 (expressing courts concern with the rule of lenity).
This rule permits the court to interpret ambiguous criminal statutes in favor of the
defendant, so that they don’t “‘penalize those whose conduct does not create the risks
of harms at which the statute aims.’” Id. See also Muscarello v. United States, 524
everything from which aid can be derived,’ … we can make ‘no more than a guess
as to what Congress intended.’ … To invoke the rule, we must conclude that there is
a ‘grievous ambiguity or uncertainty’ in the statute.’”). The court further clarifies
that the rule of lenity does not automatically permit a defendant to win. Id. See also
Precht, supra note 14, at 365 (providing “[t]he rule of lenity ‘vindicates the
fundamental principle that no citizen should be held accountable for a violation of a
statute whose commands are uncertain, or subjected to punishment that is not clearly
prescribed.’”).

96 See United States v. Nosal, 676 F.3d 854, 859 (9th Cir. 2012) (explaining their
concerns with the broad interpretation of the CFAA).

The government’s construction of the statute would expand its
scope far beyond computer hacking to criminalize any
unauthorized use of information obtained from a computer. This
would make criminals of large groups of people who would have
little reason to suspect they are committing a federal crime … we
can be properly be skeptical as to whether Congress, in 1984,
meant to criminalize conduct beyond that which is inherently
wrongful, such as breaking into a computer.
In contrast, the First, Fifth, and Eleventh Circuit Courts adopted a broader definition of authorization within the CFAA. Because this lends itself to a further array of interpretative difficulties, the respective courts leveraged their reasoning upon different underlying theories, including: agency theory, contract theory, and intended use theory. The agency theory indicates that authorization ends immediately when an employee becomes disloyal to the employer, even if the employee still technically has authorization. Meanwhile, the contract theory states that if the individual acquires information in breach of a written policy, then even the technically authorized use would fall within the premises of violating the CFAA. Finally, the intended use theory is the most broad; the courts look to the subjectivity of the company policies to determine whether authorization would technically qualify as exceeding authorized access.

97 See Din, supra note 12, at 421 (discussing that the “First, Fifth, and Eleventh Circuits have adopted broad definitions of authorization. This is largely because the addition of the CFAA’s civil liability provision has encouraged employers to increasingly use section 1030(g) to bring disloyal employees into federal court.”).

98 See Bellia, supra note 89, at 1470 (expressing issues that broad interpretation could lead to). The court would have to interpret the statute in a way that would require understanding the user’s intentions; without doing so, the CFAA could have overly broad implications and validate all states or implied restrictions on use of a system. Id. See also Din, supra note 12, at 421 (describing three types of theories the courts apply under the broader interpretation).

99 See Din, supra note 12, at 421 (stating that the duty of loyalty or agency theory provides that “authorization implicitly ends as soon as an employee becomes disloyal to his/her employer, even if he or she still has technical authorization.”). See also Bellia, supra note 89, at 1446 (categorizing “the boundaries of permissible access to a computer covered by the CFAA depend upon principles of agency law.”). “This approach typically arises in cases involving disputes between an employer and a disloyal (former) employee who uses information from the employer’s computer system to compete with the employer.” Id.

100 See Bellia, supra note 89, at 1455 (summarizing contract paradigm as policy-based, “except that the contract paradigm encompasses cases in which a court focuses on a contract as the source of restrictions on a user’s access to a computer system.”). “[A] number of factors influence the scope of the CFAA under this approach, including how carefully a court considers whether an enforceable contract exists and whether the contract addresses access to the computer system or merely use of the information the system contains.” Id. See also Din, supra note 12, at 421–22 (explaining the contract theory as finding that “if an individual acquires or utilizes information in breach of a written policy, such as a confidentiality agreement, workplace rules of conduct, or a terms-of-service agreement, then even technically authorized use constitutes unauthorized use under the CFAA.”).

101 See Din, supra note 12, at 422 (stating the “Fifth and Eleventh Circuits have both employed what they call an ‘intended use’ theory. Under this theory, courts look at
In 2021, the Supreme Court ended the longstanding circuit split and clarified the scope of “exceeds authorized access” in the landmark case *Van Buren v. United States*. The Supreme Court adopted the narrower interpretation of the CFAA through the means of a “gates-up-or-down approach.” Under *Van Buren*, an individual exceeds authorized access when they access a computer with authorization but then obtain information located in particular areas of the computer that are normally off-limits. The leading question behind the Supreme Court’s conclusion is whether the individual is entitled to access the information or not; if you need to break down a gateway or unlock permissions, then entry is a crime, if it is freely available and open, then it’s not. In adopting this narrow interpretation, the Supreme Court aimed to safeguard against incriminating millions of lawful citizens.

...
In 2022, the Supreme Court affirmed *Van Buren’s* narrow interpretation of “exceeds authorized access” in the context of data scraping from public websites in *hiQ Labs, Inc. v. LinkedIn Corp.*[^107^] hiQ, a data analytics company, scrapes data from LinkedIn users’ public profiles to create “people analytics,” which it sells to their own clients.[^108^] Their analytics can aggregate employees’ skills and identify employees who are at a high risk of being recruited from external organizations.[^109^] In 2017, LinkedIn accused hiQ of violating the CFAA; they sent hiQ a cease-and-desist letter ordering them to stop scraping its user’s data and implemented technical measures to prevent hiQ from accessing LinkedIn’s website.[^110^] On remand from the

[^107^]: See hiQ Labs, Inc. v. LinkedIn Corp., 31 F.4th 1180 (2022) (holding *Van Buren* does not apply to public websites). “[A] defining feature of public websites is that their publicly available sections lack limitations on access; instead, those sections are open to anyone with a web browser.” *Id.* “[A]pplying the ‘gates’ analogy to a computer hosting publicly available webpages, that computer has erected no gates to lift or lower in the first place.” *Id.* See Public Website definition, L. INSIDER (Jan. 29, 2023), archived at https://perma.cc/34W2-KURH (providing various definitions for a public website). “Public website means that portion of the Website which is accessible to the general public, without use of credentials and/or authentication, as applicable.” *Id.*

[^108^]: See hiQ Labs, Inc., 31 F.4th 2 at 1187 (stating HiQ was founded in 2012). HiQ “scrapes information that LinkedIn users have included on public LinkedIn profiles, including name, job title, work history, and skills. It then uses that information, along with a proprietary predictive algorithm, to yield ‘people analytics[.]’” *Id.* See also Mark Weidick, 2018 Pharma Talent Report: Focus on Immuno-Oncology Professionals, hiQ (Nov. 17, 2022), archived at https://perma.cc/7KRP-DC95 (characterizing hiQ products as “insights designed to help companies understand their greatest asset … their people.”).

[^109^]: hiQ Labs, Inc., 31 F.4th at 1187 (2022) (describing two types of analytics, Keeper and Skill Mapper, that hiQ offers for products). See also Weidick, supra note 108 (finding for their client that there was urgency in their investments of Immuno-Oncology professionals). “The findings in this report should elicit a sense of urgency … [I-O talent] are nested in a dynamic environment wherein the average employee is switching jobs about every 4 years.” *Id.*

[^110^]: See Reena Bajowala et al., Narrowing the Scope of the Computer Fraud and Abuse Act: Ninth Circuit Finds in Favor of Data Aggregator Scraping Data from Public Website, 8 Pratt’s Priv. & Cybersecurity L. Rep., 237 (2022) (summarizing that “[i]n 2017, LinkedIn sent hiQ a cease-and-desist letter demanding that it stop scraping data from LinkedIn and accusing hiQ of violating, among other things, the CFAA and LinkedIn’s User Agreement by collecting users’ data. LinkedIn also implemented technical measures[,]”). See also Christopher Burgess, HiQ v LinkedIn court ruling will have a material effect on privacy, CSO (June 1, 2022), archived at https://perma.cc/C4QM-BWEF (describing relationship between the companies and hiQ’s business reliance on LinkedIn). “hiQ was able to get an injunction against LinkedIn’s actions and remain afloat” while the case was pending. *Id.*
Supreme Court, the Ninth Circuit held that “‘[a] violation of the terms of use of a website–without more–cannot establish liability under the CFAA.’” The court noted that the CFAA is meant to be an “anti-intrusion” statute rather than a “misappropriation” statute, so in order to violate the CFAA an authentication requirement, like a password gate, would need to have been broken. Ultimately, if a user accesses publicly available data on a network that permits public access to data, like hiQ did from LinkedIn, this does not meet the threshold of “without authorization” under the CFAA.

2. Pros and Cons of Legal Data Scraping for Companies

Many may consider the rulings of Van Buren and hiQ as major breakthroughs for analytic companies or other businesses looking to reap the benefits of data scraping. In fact, there are countless reasons why businesses may look to data scraping techniques to remain modern in today’s tech-driven world. Data scraping is a relatively cheap and fast way for businesses to generate information about their current and future landscape; companies can leverage their

---

111 See hiQ Labs, Inc., 31 F.4th at 1195-6 (2022) (quoting precedent of Nosal). “The CFAA was enacted to prevent intentional intrusion onto someone else’s computer–specifically, computer hacking.” Id.
112 See Bajowala et al., supra note 110 (describing the CFAA as an anti-intrusion statute rather than a misappropriation statute). “[T]o violate the CFAA’s ‘without authorization’ provision, ‘an authentication requirement, such as a password gate, is needed to create the necessary barrier that divides open spaces from closed spaces on the Web.’” Id.
113 See hiQ Labs, Inc., 31 F.4th at 1201 (2022) (rationalizing that “when a computer network generally permits public access to its data, a user’s accessing that publicly available data will not constitute access without authorization under the CFAA.”). “The data hiQ seeks to access is not owned by LinkedIn and has not been demarcated by LinkedIn as private using such an authorization system.” Id. See also W. Connor McRory, Let the Bots be Bots: Why the CFAA Must Be Clarified to Prevent the Selective Banning of Data Collection Facilitating Private Social Media Information Monopolization, 16 BROOK. J. CORP. FIN. & COM. L. REV. 279, 302–03 (Dec. 1, 2021) (suggesting that there are different levels to accessing publicly available data).
114 See Mackey & Opshal, supra note 90 (describing the decisions as a victory for the internet and the Van Buren decision as “especially good news for security researchers, whose work discovering security vulnerabilities is vital [and] … often requires accessing computers in ways that contravene terms of service.”).
115 See Richard Sutherland, What is web scraping, and how are businesses using it to gain an edge?, TECHRADAR.PRO (Nov. 17, 2022), archived at https://perma.cc/4MWX-B7AL (explaining how businesses use web scraping to gain competitive edge).
internal resources to data scrape themselves or contract with more sophisticated third-party providers. For example, Smartproxy offers large-scale, customizable, data scraping plans that cost anywhere from $50.00 per month for 25,000 requests to $500.00 per month for 625,000 requests. For these prices, the benefits are vast; data scraping can generate leads, optimize pricing strategies from competitors, assist in transitioning websites, and generally help a business to better understand their customer base.

On the other side, the climate of allowing legal scraping of publicly available information can make it quite challenging for companies whose business models rely on hosting public information, like LinkedIn and Meta. Social media applications and networking...
sites are often free for users in exchange for the collection of their data.\textsuperscript{120} But, if this data is public, how are hosting sites able to prevent third parties from collecting the same data?\textsuperscript{121} Companies may have to consider implementing preventative measures that fall within the realm of the CFAA, such as implementing controls and password gates for certain parts of their sites, in order to curb outsiders from accessing this publicly available information.\textsuperscript{122} For example, Facebook’s help center references that they have a dedicated External Data Misuse team to make scraping more difficult, but acknowledges they will “never be able to fully prevent all scraping without harming people’s ability to use [their] apps and websites the way they enjoy.”\textsuperscript{123} Organizations are at the crossroads of an ethical and moral dilemma; they must consider whether to restructure their business models to protect user data or risk exposure of third-party data scraping.\textsuperscript{124}

\textsuperscript{120} See Swish Goswami, \textit{What The Future Of Consumer Data Ownership Looks Like}, FORBES (Nov. 23, 2021), archived at https://perma.cc/32Z9-V5DD (indicating users were typically fine with sharing their data with platforms like Facebook, Twitter and YouTube, when they were created because that meant they could use the sites for free).

\textsuperscript{121} See Crook, \textit{supra} note 119 (suggesting that from the companies perspective “the Ninth Circuit’s opinion . . . makes it harder to prevent third parties from collecting and monetizing that information.”).

\textsuperscript{122} See Crook, \textit{supra} note 119 (stressing that some “companies may have to consider whether they require the equivalent of ‘gates down’ approaches going forward if they seek to invoke the protections of the CFAA.”). See also Erez Hasson, \textit{Web Scraping: The Fine Line Between Business Intelligence and Data Privacy Violation. Is it Legal?}, IMPERVA (May 19, 2021), archived at https://perma.cc/NZ2W-UA3L (stating web scraping “remains an issue that is tricky to solve in a legal manner” so organizations are likely to adopt preventative measures).

\textsuperscript{123} See \textit{What is data scraping and what can I do to protect my information on Facebook?}, FACEBOOK (Nov. 18, 2022), archived at https://perma.cc/QB7F-PJM9 (describing methods Facebook takes to prevent unauthorized scraping). Facebook has a dedicated EDM team “that works to make it harder and more costly for scrapers to gather data from [their] . . . services or profit from it.” \textit{Id.} Facebook makes it clear that they will not be able to prevent 100\% of scraping. \textit{Id.}

\textsuperscript{124} See Hasson, \textit{supra} note 122 (indicating there is a fine line between business intelligence and data privacy). Web scraping “poses a moral dilemma for organizations. As more of them realize that not leveraging certain techniques may place them at a disadvantage, the probability of them turning to said techniques is high . . . considering no firm legal action is being taken to put a halt to web scraping operations.” \textit{Id.} See also Burgess, \textit{supra} note 110 (discussing the negative implications of what could happen if LinkedIn blocked the public access option on their site). If LinkedIn wanted to restrict public access, they could eliminate the “‘public access option, albeit at a cost to the performance of many users and, possibly, to its own bottom line.’” \textit{Id.}
3. The Privacy Imperative Data Scraping Poses for Consumers

Consumers are recognizing the value their data holds and how easily it can be collected, sold, and used without their knowledge.\(^\text{125}\) The consumer information extracted from data scraping is personal; it can consist of phone numbers, email addresses, physical addresses, social security numbers, financial information, and purchase behaviors.\(^\text{126}\) Though the landscape is beginning to change, consumers are not always aware of what type of data is being aggregated or how securely it is being done.\(^\text{127}\) Given that in 2020, the average consumer was associated with 8.4 social media accounts, the opportunity for third parties to scrape their consumer data is amplified, leaving consumers at greater risk of cyber security threats and data breaches.\(^\text{128}\) Due to the increase in reported data breaches, consumer trust in certain companies and industries is declining.\(^\text{129}\) Ironically, consumers are becoming more intentional with who they share their data with, but are

\(^{125}\) See Goswami, supra note 120 (highlighting consumers realize the value of their data, “but they’re also becoming privy to the invasive nature in which their data is collected and deployed.”). See also Anant, supra note 8 (citing high-profile consumer-data breaches as why consumers are lacking trust when sharing with organizations).

\(^{126}\) See Ben Canner, The Dangers of Data Scraping: Do You Know What’s Out There?, S.O.L.'s REV. (Aug. 24, 2020), archived at https://perma.cc/H9V3-5BNE (opining that data scraping “extracts human data, such as email addresses, phone numbers, shopping behaviors and more.”).

\(^{127}\) See id. (referencing the problems on both sides of the data collection interaction). On the scraped users’ side, they often don’t know what information is being collected or that someone is aggregating their data in the first place. Meanwhile, scrapers may not configure the databases of collected information or secure them at all. The latter allows hackers of all calibers to access critical consumer and employee data.

\(^{128}\) See Brian Dean, Social Network Usage & Growth Statistics: How Many People Use Social Media in 2022?, BACKLINKO (Oct. 10, 2021), archived at https://perma.cc/6XZD-CZSZ (finding key statistics related to social media users). “The average number of social media accounts [was] 8.4 per person in 2020.” Id. Social media growth rates are only going to increase year after year. Id.

\(^{129}\) See Anant, supra note 8 (indicating that consumer trust levels are generally low but the industry has an affect). The healthcare and financial services industries received the highest rating, 44% for trust; this is likely due to the volume of sensitive and personal information being processed within these industries. Id.
not adequately taking protective precautions to keep their data secure.\textsuperscript{130}

\textbf{IV. Analysis}

\textit{A. What Legal Data Scraping Means For The Consumer}

In the aftermath of \textit{hiQ}, it is imperative that consumers not only understand how companies are using their data, but also that they are empowered to take control over how their personal data will be used in the future.\textsuperscript{131} The Ninth Circuit acknowledges that giving companies free rein on how to control their users’ data could create information monopolies that disserve the public interest.\textsuperscript{132} The court’s opinion makes it quite difficult for companies to prevent third parties from obtaining data stored publicly on their website.\textsuperscript{133} Though

\textsuperscript{130} See \textit{id.} (highlighting consumer empowerment and corresponding actions to protect data). “[C]onsumers often want to restrict the types of data that they share with businesses.” \textit{Id.} Consumers generally have greater control to do so with options like built-in cookie blockers, ad-blocking software and incognito browsers. \textit{Id.} Consumers may opt out of sharing data with transactions they deem less important and might even walk away from conducting certain business if they don’t feel comfortable with certain privacy policies. \textit{Id.} Consumers generally aren’t taking adequate protections to protect their data; only 41\% of respondents have set their browser to disable cookies, and only 18\% have masked their identity online. \textit{Id.} See also Goswami, supra note 120 (explaining what the future of consumer data privacy could look like). Specifically, the government could crack down on the level of tracking from big tech and consumers could be awarded for their data through varying business models. \textit{Id.}

\textsuperscript{131} See \textit{hiQ Labs, Inc. v. LinkedIn Corp.}, 31 F.4th 1180 (2022) (holding \textit{hiQ’s} access to \textit{LinkedIn} users’ public data does not violate the CFAA).

It is likely that when a computer network generally permits public access to its data, a user’s accessing that publicly available data will not constitute access without authorization under the CFAA. The data \textit{hiQ} seeks to access is not owned by \textit{LinkedIn} and has not been demarcated by \textit{LinkedIn} as private.\textit{[.]} \textit{Id.} The court acknowledges that “giving companies like \textit{LinkedIn free rein . . . [on]} who can collect and use data . . . that the companies do not own, that they otherwise make publicly available to viewers . . . risks the possible creation of information monopolies that would disserve public interest.” \textit{Id.}

\textsuperscript{132} See \textit{id.} (opining on the dangerous implications this holding could have). “[G]iving companies like \textit{LinkedIn free rein . . . [on]} who can collect and use data . . . that the companies do not own, that they otherwise make publicly available to viewers . . . risks the possible creation of information monopolies that would disserve public interest.” \textit{Id.}

\textsuperscript{133} See Crook, supra note 119 (stating that the court’s opinion makes it more difficult to prevent third parties from collecting and monetizing information). The facts of
there are changes businesses should be implementing to safeguard consumers’ personal information, the first change lies in the hands of the consumer.  

Consumers, as the owners of their data, should be the primary decision-makers of how their data is being stored, transmitted, and sold by businesses. Those who are wary of data privacy can initiate control in a number of ways, from reading the fine print to using a virtual private network, or even reconsidering their use of free applications. For example, the consumer can implement simple search capabilities by highlighting keywords that are typically

hiQ highlight the importance of the public accessibility of the data and whether the company asserts any control over the data. Id. 

134 See Bajowala, supra note 110 (highlighting that companies seeking to protect publicly available online information should “evaluate the applicability of the CFAA to their websites”). If only the company’s terms of service are violated by a third-party, they may not have any protection under the CFAA at all. Id. “Whether a company is seeking the protection of the CFAA is a data aggregator, an operator of a public social networking website, or otherwise allows public access to company data on the internet, the hiQ opinion suggests that it should carefully analyze its data access policies and practices.” Id. See also Freedman, supra note 5 (evaluating that consumers are naïve in respect to their data protection efforts). A 2022 Ipsos poll revealed that “only 34% of Americans think that companies adequately safeguard [their data], [and] … [o]nly 16% of respondents took all six data security measures about which Ipsos asked.” Id.

135 See Canner, supra note 126 (explaining that data scraping poses issues for unaware consumers). “On the scraped users’ side, they often don’t know what information is being collected or that someone is aggregating their data in the first place.” Id. See Goswami, supra note 5 (stating “[y]ou own your data, and as a result, you should decide how and when your data is used and be compensated for it appropriately.”). See also Rahnama, supra note 9 (explaining the basic principle that “personal data is an asset held by the people who generate it.”).

136 See Guynn, supra note 3 (outlining the disconnect that users have with their data, specifically with user agreements and terms of service policies). “We needlessly put ourselves at risk by signing away all kinds of rights over what personal data an app or website collects, how they use it, with whom they share it and how long they keep it[.]” Id. A Deloitte survey revealed that 91% of people consent to terms of service without reading it. Id. Certain aspects of these documents are important to read, like “granting a company the right to sell your personal information to third parties, trace your movements using GPS and other tracking capabilities, harvest your device identifiers or track your device’s IP address[.]” Id. See also Freedman, supra note 5 (listing options consumers have to keep their data private). By blocking ads and trackers, businesses can’t collect or track your data. Id. “When you browse the internet with a VPN, you tunnel information from your device to a server. This tunneling hides your browsing activity, putting up a wall between businesses and your data.” Id. Additionally, you can sign up for non-essential services with fake information to protect your data and avoid linking apps so that they are not able to share data between one another. Id.
associated with personal information and data retrieval methods, such as “accept,” “authorize,” and “opt-out”; consumers can also enable unique browser extensions that automatically scan and rate user agreements for them. Additionally, companies like FastMail and Zoho offer alternatives to Gmail and Google Docs in exchange for a fee and peace of mind that their data will not be monetized. Inevitably, each consumer will have varying levels of comfort and flexibility with respect to the use and exposure of their data; thus, each consumer should be the driver of their personalized experience.

B. How Big Business Can Maintain Customer’s Confidence in Utilizing Their Data

Based on the level of mistrust that consumers have towards companies handling their personal data, there is a significant opportunity for businesses to differentiate themselves by prioritizing consumer data privacy. Right now, a majority of organizations’

137 See Guynn, supra note 3 (explaining that searching for keywords or phrases in documents will direct the user to which sections of the terms and conditions that users should be aware of). Users should be more cautious in sections that say you must ‘accept,’ ‘agree’ or ‘authorize.’ Id. Furthermore, ‘third parties,’ ‘advertising partners,’ and ‘affiliates,’ are all key phrases to look out for. Id. The words ‘retain’ or ‘retention’ will notify users how long a company keeps their personal information and ‘opt out’ will direct users on how to turn off the company’s ability to collect your personal information. Id.

138 See Glikman, supra note 4 (finding that many users who oppose corporations using their personal data are not willing to pay an extra fee to protect their own data). There are a number of services available that offer to protect our data, but “[t]he dominant model remains one of free access to services in exchange for a commercial use of our data.” Id. Alternatives are often fee-based; FastMail offers an alternative to Gmail and Zoho offers an alternative to Google Docs. Id.

139 See Goswami, supra note 120 (opining on how consumer ownership could change in the future). “If consumers own their data and are in control of how it is used, then everyone is happy.” It is their choice whether a consumer wants their information to be private or available to brands; if a consumer is not as concerned with privacy, then they should have the ability to exchange their data for a value, “the level of which they are willing to accept in exchange for their data is up to each individual consumer.” Id.

140 See Anant, supra note 8 (revealing research shows that consumers do not trust companies to handle their data or protect their privacy). “[C]onsumers respond to companies that treat their personal data as carefully as they do themselves.” Id. Companies can differentiate themselves by taking “deliberate, positive measures” to protect consumer data. Id. Consumers general trust is at an overall low in most industries, with the exception of healthcare and financial services. Id. See also Rahnama, supra note 9 (outlining the new rules of data). “If your organization
information technology budgets are spent trying to manage the internal complexities of data handling, rather than enhancing consumer trust and improving their overall customer experience.\textsuperscript{141} In order to reap higher levels of consumer trust, organizations must be willing to provide increased transparency to current and potential customers.\textsuperscript{142} This can take place in a number of formats; businesses can update operating agreements to reflect consent in layman’s terms, minimize data mapping and collection to strict requirement standards, update data-storage policies, limit access, and modernize internal data infrastructure.\textsuperscript{143} For example, in 2021, Apple updated the iOS generates any value from personal data, you will need to change the way you acquire it, share it, protect it and profit from it.” \textit{Id.}\textsuperscript{141} See Rahnama, \textit{supra} note 9 (describing the complex internal data collection operations of organizations). “[U]p to 90 percent of current IT budgets are spent simply trying to manage internal complexities, with precious little money actually spent on data innovation that improves either productivity or the customer experience.” \textit{Id.} \textit{See also} Brashear, \textit{supra} note 87 (describing the outdated business practices with respect to data). “Businesses often rely on a mix of legacy technologies, bespoke applications, and modern as-a-service offerings. This can lead to a technical workforce that spends more time operating and maintaining a complex patchwork of technologies than in valuable activities involving data-driven innovation.” \textit{Id.}\textsuperscript{142} See Rahnama, \textit{supra} note 9 (providing that “[f]irms need to consistently cultivate trust with customers, explaining in common-sense terms how their data is being used and what’s in it for them.”). \textit{See also} Anant, \textit{supra} note 8 (explaining customer-facing best practices). “It is important for organizations to communicate transparently: customers should know when and why their data are being collected. Many companies are adding consumer privacy to their value propositions and carefully crafting the messages in their privacy policies and cookie notices to align with the overall brand.” \textit{Id.} \textit{See also} Brashear, \textit{supra} note 87 (arguing the “groundwork for using data as a strategic asset is building consensus for change in processes, technologies, and those who employ them.”).\textsuperscript{143} See Anant, \textit{supra} note 8 (explaining that data maps or registers categorize types of data companies collect from consumers). “Companies need to know which data they actually require to serve customers. Much of the data that is collected is not used for analytics and will not be needed in the future. Companies will mitigate risk by collecting only the data they will probably need.” \textit{Id.} A great approach to update data-storage and data-security policies is to account for all of the different categories of data, as each category may require unique mechanisms of storage. \textit{Id.} Organizations should follow industry best-practices by storing data in a limited number of systems, depending on data type or classification; “[a] smaller systems footprint reduces the chance of breaches.” \textit{Id.} \textit{See also} Brashear, \textit{supra} note 87 (outlining results of a 2021 Deloitte survey). “40% [of CFOs] report[ed] ‘low’ or ‘medium’ level of sophistication across a range of data priorities [including infrastructure].” \textit{Id.} \textit{See Justices Clarify Scope of Anti-Hacking Law, supra} note 20 (stating that access to sensitive information needs to be limited “by policy, contract, and technology—to only those who need the information to perform their
operating system to enable users the opportunity to turn off cross-app tracking on their mobile devices.\footnote{Apple is leading the market in terms of their privacy protections, resulting in the attraction of new customers and a significant bottom line impact for mobile applications that profited from cross-app tracking. As a result, it is likely that other companies will begin to operate under similar business models or offer equivalent incentives for consumers to choose their product over that of a competitor.}

With the increase of consumer personalization and control over their data, there will be a greater value placed on consumer data that has been consented to. Organizations are going to need consumers’ explicit approval and incentivizing them could become an essential component for organizations to maintain their competitive data-driven edge. There is a relatively large number of consumers in the functions.

\footnote{See Rahnama, supra note 9 (providing that “firms need to consistently cultivate trust with [their] customers, explaining in common-sense terms how their data is being used and what’s in it for them.”).}

\footnote{See id. (describing how Apple provided customers with power and agency over their data).}

\footnote{See Rahnama, supra note 9 (describing how Apple’s upgrade to its iPhone operating system allowed users to shut down data harvesters’ ability to track them across their many apps . . . [i]t also bit hard into companies that rely on cross-app tracking . . .).}

\footnote{See id. (explaining the financial impact cross-app tracking has on social media giants). As a result of Apple’s privacy update, “it cost the major social media sites $10 billion in lost revenue in the second half of 2021.” Id. Meta expects that it will cost them roughly another $10 billion in 2022. Id. “Apple has made privacy protection a market differentiator: device manufacturers and app developers now use privacy features to draw new users.” Id.}

\footnote{See Rahnama, supra note 9 (referencing the clear endpoint where individuals will be able to exercise full control over their data). “While consumers still seek the conveniences and benefits that flow from their data, they will be the ones to set the terms over what data they share and who they share it with.” Id. Consumers want the protection from the government. Id. It is only a matter of time until competition over data privacy will impact financial bottom lines. Id.}

\footnote{See Rahnama, supra note 9 (alleging that data collected with “meaningful consent will soon be the most valuable data of all, because that’s the only data companies will be permitted to act upon.”). Firms will likely follow the lead of other data cooperatives and sharing services by securing each user’s consent for the option they align most comfortable with. Id. “The more robust and thorough your consent practices are, the more valuable your customer database becomes.” Id. See also Busby, supra note 88 (contesting that consumer data is at the top of the list for companies wishing to extract value from). “Customer data that feeds new value propositions, new and improved experiences, and new revenue models is how winners will distance themselves from . . . [competitors].” Id.}

\footnote{See Goswami, supra note 120 (distinguishing the younger generations response to the growing economy of data and privacy). It is highly likely that younger
marketplace that are willing to let companies collect their information in exchange for monetization or something else of value.\textsuperscript{149} Thus, it is critical that companies understand how consumers value different categories of data.\textsuperscript{150} A Harvard Business Review study found that consumers consider “self-reported data,” such as e-mail addresses or employment history, to have a lower value than “profiling data” that is sold to third parties.\textsuperscript{151} Generally, consumers attribute value to enhanced products and services; therefore, consumers are willing to share their data in exchange for convenience or privilege.\textsuperscript{152} For example, Disney uses profiling data from its MagicBand bracelet to enhance customers experience; they can unlock their hotel rooms, charge food and merchandise, and check in at reservations by simply

consumers’ feelings towards the loss of privacy due to “data-mining and scraping by tech companies, brands, and marketers is: ‘What’s in it for me?’” \textit{Id. See also} Freedman, \textit{supra} note 5 (finding that “78\% of respondents said they wanted to require companies to obtain their consent before accessing and using their [personal] data.”).

\textsuperscript{149} See Goswami, \textit{supra} note 120 (characterizing younger consumers as being okay with their data being collected so long as they receive something of value in exchange for it).

\textsuperscript{150} See Morey, \textit{supra} note 4 (referencing that “[i]f companies understand how much data is worth to consumers, they can offer commensurate value in return for it.”).

\textsuperscript{151} See Morey, \textit{supra} note 4 (describing different types of data and corresponding values reported to each category). Self-reported data is “information people volunteer about themselves, such as their e-mail addresses, work and educational history, and age and gender[.]” \textit{Id.} Digital exhaust data consists of “location data and browsing history, which is created when using mobile devices, web services, or other connected technologies[.]”. \textit{Id.} Profiling data consists of “personal profiles used to make predictions about individuals’ interests and behaviors, which are derived by combining self-reported, digital exhaust, and other data.” \textit{Id.} People value self-reported data the least, digital exhaust data a little more, and profiling data the most. \textit{Id.}

\textsuperscript{152} See Morey, \textit{supra} note 4 (describing the value consumers see in sharing their data). “Users hand over a lot of data, but they get convenience and a sense of privileged access in return, making the trade-off worthwhile. Consumers know exactly what they’re signing on for, because Disney clearly spells out its data collection policies in its online MagicBand registration process . . .” \textit{Id. See also} Rea, \textit{supra} note 4 (summarizing Netflix’s approach to leveraging data). Netflix uses individual subscriber data to tailor its content offerings; this is visible whenever you let a family member use your Netflix account and they proceed to watch shows that you would never consider. \textit{Id.} When other users use the account, the recommendations begin to change; “[t]his use of granular data to customise [sic] the individual user experience makes subscribers sticky and reduces the costly risk of producing or commissioning unpopular content.” \textit{Id.}
holding up their MagicBand to sensors around the parks.\textsuperscript{153} It is a cyclical give-and-take approach – the greater transparency businesses provide, the greater trust consumers will have; the more trust consumers have, the more confident they will feel sharing their data.\textsuperscript{154}

C. Need for Nationwide Privacy and Security Laws

The current patchwork of state-created privacy laws in the United States pose a considerable challenge for businesses to identify whether they fall in the scope of said regulations or whether they fall short of protecting consumers.\textsuperscript{155} The idea of increased transparency for consumers is quite relevant in the regulatory landscape of privacy; consumers often believe their privacy is protected, until it’s too late and their own data becomes exposed.\textsuperscript{156} If companies have difficulty navigating the variations of state privacy laws, then it is inevitable that consumers would also have difficulty in understanding how their data

\begin{flushleft}
\textsuperscript{153} See Morey, supra note 4 (explaining how Disney using profiling data to enhance customers’ experiences). Disney uses profiling data “gathered by its MagicBand bracelet to enhance customers’ theme park and hotel experiences and create targeted marketing. By holding the MagicBand up to sensors around Disney facilities, wearers can access parks, check in at reserved attractions, unlock their hotel doors, and charge food and merchandise.” \textit{Id.}

\textsuperscript{154} See \textit{id.} (finding correlations between value of data and exchange of services).

[\textit{T}he value consumers place on their data rises as its sensitivity and breadth increase from basic information that is voluntarily shared to detailed information about the consumer that the firm derives through analytics, and as its uses go from principally benefitting the consumer (in the form of product improvements) to principally benefitting the firm (in the form of revenues from selling data).]

\textit{Id.}

\textsuperscript{155} See Klosowski, supra note 47 (indicating there is a lack of privacy laws in the United States). “Currently, privacy laws are a cluttered mess of different sectoral rules.” \textit{Id.} Amie Stepanovich, executive director at the Silicon Flatirons Center at Colorado Law states that the US has a bunch of disparate federal and state laws. \textit{Id.} “The United States doesn’t have a singular law that covers the privacy of all types of data.” \textit{Id.} See also Data Privacy Laws by State, supra note 51 (suggesting that the subtle differences in U.S. state laws have the potential to confuse even the most seasoned compliance professionals).

\textsuperscript{156} See Klosowski, supra note 47 (recognizing “[m]ost people believe they’re protected, until they’re not[,]”). Former Chief Technologist of the Federal Trade Commission states that “this ecosystem is primarily hidden from view and not transparent, [so] consumers aren’t able to see and understand the flow of information.”\textit{Id.}
is legally protected. Currently, California’s CCPA is the strictest privacy law in the nation, but its protection is limited to: (1) entities that do business in California, (2) companies who satisfy annual revenue thresholds of $25 million, (3) companies that process personal data of more than 50,000 individuals, and (4) companies that acquire 50% of their revenue from selling data. Among the other few states that have variations of privacy laws enacted, some are too business-model friendly. For example, Virginia’s Consumer Data Protection Act was written into legislation with a strong influence by Amazon; other state bills, like Connecticut, Florida, and Oklahoma originally failed because they included private rights of actions. The different variations of state privacy laws may in theory sound like progress has been made for the United States, but they have proven to be ineffective

---

157 See Klosowski, supra note 47 (noting the risk of too many state laws generating confusion for consumers). See also Harrington, supra note 21 (providing that U.S. privacy laws vary). Although many of the U.S. privacy laws share some of the key provisions, like obtaining consumer consent, there are crucial differences between the requirements of each law. Id. Some laws provide that businesses who violate the law will be required to pay certain penalties or fines, and in a handful of states a consumer themselves has the right to sue. Id. See also Federal Cybersecurity and Data Privacy Laws Directory, supra note 22 (stressing that the variations in state cybersecurity and data breach notification laws pose a considerable challenge for organizations that operate in all 50 states).

158 See Harrington, supra note 21 (outlining the difference in the CCPA versus GDPR regulation). The “CCPA only covers entities that do business in California. This regulation applies to entities satisfying thresholds such as annual revenues above $25 million, any organization that processes personal data of more than 50,000 individuals, and those entities that acquire 50 percent of their revenue from selling data.” Id.

159 See Klosowski, supra note 47 (mapping that only three states have comprehensive data privacy laws).

160 See id. (stating Virginia’s Consumer Data Protection act was written with strong input from Amazon). VCDPA is considered a weak bill by many; “It is based on opt-out consent. There are no civil-rights protections. There is no private right of action. A lot of the provisions are business-model affirming. It essentially allows big data-gathering companies to continue doing what they have been doing.” Id. Hayley Tsukayama, a legislative activist, states that it is an insult to public injury to underfund the public enforcement, “especially in those states where they don’t allow a private right [to sue]. . . .” Id. “California’s law has a limited private right of action related to negligence with regard to a data breach. The Colorado and Virginia laws don’t even have that. Several bills, including those in Connecticut, Florida, Oklahoma, and Washington, failed to become laws because they included a private right of action.” Id.
in practice; United States consumers need the protection of a federal data privacy law.161

1. Proposal For A Federal U.S. Privacy Law

An overarching national privacy regulation, like the European Union’s GDPR, would streamline the compliance process for companies and provide greater transparency to consumers.162 Because GDPR is the most strict and far-reaching privacy regulation in the world, the United States should implement a privacy regulation that runs parallel to the GDPR rather than interfere with it.163 Less regulated and domestic-only industries may feel some disruption to their current operating models, but major industries, like financial services and technology, are already accustomed to more strict, or GDPR-compliant, standards.164 Like GDPR, a United States Privacy Law should not have a jurisdictional threshold limited to the borders of the United States; rather, it should apply to any company that processes the personal data of United States citizens, regardless of revenue and processing threshold amounts.165

161 See Klosowski, supra note 47 (opining on the need for a federal data privacy law). Whitney Merrill, a privacy attorney and data protection officer, states that “[w]e need a federal law that thinks about things in a much more consistent approach … to make sure that consumers understand and have the right expectation over rights that they have in their data.” Id. Senior legislative counsel, Kate Ruane, stressed that “[c]onsumer data privacy laws can give individuals rights to control their data, but if poorly implemented such laws could also maintain the status quo.” Id.
162 See Anant, supra note 8 (highlighting that privacy regulations are evolving to protect consumers). “The GDPR gives consumers easier access to data that companies hold about them and makes it easier for them to ask companies to delete their data.” Id.
163 See Anant, supra note 8 (explaining how other countries outside of the EU are implementing their own data-privacy regulations). In August 2020, Brazil’s General Data Protection Law went into effect, known as “LGPD.” Id. “LGPD is an overarching, nationwide law centralizing and codifying rules governing the collection, use, processing, and storage of personal data.” Id.
164 See Anant, supra note 8 (outlining the compliance investments companies have already had to make as a result of GDPR). “In total, Fortune Global 500 companies had spent $7.8 billion by 2018 preparing for GDPR, according to an estimate by the International Association of Privacy Professionals.” Id. Companies have hired data-protection officers, as required by the GDPR for certain companies. Id. For example, Microsoft is applying CCPA requirements to all US citizens, rather than just their California customers; companies are using the most restrictive standards for their own legal requirements to make them fool-proof. Id.
165 See Data Privacy Laws by State, supra note 51 (comparing GDPR to different state-level privacy laws in the U.S.).
The GDPR contains critical regulatory provisions around data security, data processing, and data consent that the United States should replicate in its own privacy law. Recital 78 of GDPR describes appropriate technical and organizational measures that companies must adhere to; these include internal procedures such as pseudonymizing personal data, allowing the data subject to monitor the data processing, limiting internal access to personal data, and implementing technical measures like two-factor authentication and encryption. Next, data processing in the United States should only be permissible under a similar set of circumstances to Article 6 of the GDPR; there must be a legitimate lawful basis to process the data, including situations where there is unambiguous consent, a background check or other legal obligation is involved, or the data is relevant to preventative healthcare and emergency situations. Additionally, the United States Privacy Law should look to Article 7

GDPR requires compliance by any entity that processes personal data in the context of activities of an “establishment” in the EU, or processes personal data of individuals in the EU related to the offering of goods and services to them or monitoring their behavior. There is no revenue threshold, processing threshold, or broker threshold.

Id. Meanwhile, the CCPA only applies to entities that do business in California and meet specific revenue and data processing thresholds. Id. See What is GDPR, the EU’s new data protection law?, supra note 61 (highlighting the key regulatory points of the GDPR). There are seven different data protection principles of the GDPR. Id. The GDPR requires that data be handled securely by implementing “appropriate technical and organizational measures.” Id. Furthermore, every default action that an organization takes must consider data protection, “by design and by default.” Id.

See Recital 78, supra note 63 (listing the appropriate technical and organizational measures required by the GDPR). In order to demonstrate compliance, companies should adopt and implement internal data protection measures. Id. “Such measures could consist, inter alia, of minimizing the processing of personal data, pseudonymizing personal data as soon as possible, transparency with regard to the functions and processing of personal data, enabling the data subject to monitor the data processing, enabling the controller to create and improve security features.” Id.

See What is GDPR, the EU’s new data protection law?, supra note 61 (listing instances where it is permissible to process data under the GDPR). There must be a lawful justification to process someone’s personal data under the GDPR. Id. If a consumer has opted to receive your email marketing, this is considered to be unambiguous consent to process their data. Id. Data may need to be processed to enter into a contract or to comply with another legal obligation. Id. Data may need to be processed to save lives or to perform an official public function. Id. The most flexible basis of the GDPR is if the entity has a legitimate interest to process someone’s personal data, but there are a variety of factors that need to be considered here. Id.
of the GDPR to define what constitutes consent.\textsuperscript{169} Consent should be specific, informed, and revocable; it should be documented in clear and plain language and the process to withdraw consent should be as simple as providing consent.\textsuperscript{170} Finally, the penalties for non-compliance should mirror that of the GDPR’s, so that they reflect a percentage of the company’s operating profits, instead of a capped-amount per violation like the CCPA; the severity of impact on a company’s bottom line will incentivize companies to comply.\textsuperscript{171} An overarching federal regulation will fill the void for consumers who currently lack any sort of state-level privacy protections, but more importantly, it will provide a uniform level of privacy protection for all U.S. citizens.\textsuperscript{172}

\begin{flushleft}
\textsuperscript{169} See What is GDPR, the EU’s new data protection law?, supra note 61 (providing rules on what constitutes consent from a data subject). Consent must “freely given, specific, informed and unambiguous.” Id.

\textsuperscript{170} See id. (stating that requests for consent must be in plain language and the withdraw of consent must be simple). “Data subjects can withdraw previously given consent whenever they want, and you have to honor their decision. You can’t simply change the legal basis of the processing to one of the other justifications.” Id.

\textsuperscript{171} See Data Privacy Laws by State, supra note 51 (outlining consequences of non-compliance for GDPR and CCPA). “The consequences of non-compliance of GDPR are administrative fines up to €20 million or 4% of total worldwide annual turnover of the preceding financial year, whichever is higher.” Id. Violators of the CCPA face “civil penalties of up to $7,500 per intentional violation or $2,500 per unintentional violation. In actions brought by consumers for security breach violations, the consequences are statutory damages not less than $100 and not greater than $750 per consumer per incident or actual damages, whichever is greater.” Id. See also Anant, supra note 8 (explaining that Brazil’s LGPD has formidable fines). “While the fines are less steep than the GDPR’s … failing to comply with the LGPD could cost companies up to 2 percent of their Brazilian revenues.” Id.

\textsuperscript{172} See Klosowski, supra note 47 (asserting that with the wide range of laws people get confused about what rights they do and do not have). According to privacy experts, four areas deserve basic protections. Id.

[1] Data collection and sharing rights: Laws should give people the right to see what data various companies have collected on them, to request that companies delete any data they’ve collected, and to take data easily from one service to another. This also includes the right to tell companies not to sell (or share) your data to third parties. . . . [2] Opt-in consent: A company should have to ask you if it may share or sell your data to third parties. You shouldn’t have to spend hours opting out of the collection of your private data through every service you use. [3] Data minimization: A company should collect only what it needs to provide the service you’re using. [4] Nondiscrimination and no data-use discrimination: A company shouldn’t discriminate against people who exercise their privacy rights[.]
\end{flushleft}
2. Proposal For Amendments To The CFAA

The holdings of *Van Buren* and *hiQ* shed light onto a vulnerability within the CFAA. Accessing publicly available information on a website without trespassing through a password gate, or other barrier of entry, does not equate to “without authorization” liability under the CFAA; companies whose business models rely on data scraping will continue to profit on users’ data so long as they fall within this loophole. Though businesses should be adjusting their terms of service and user agreements, as noted above, the CFAA needs to provide a more protective consumer-focused default rule. Perhaps this involves differentiating between the different types of public websites and the different thresholds of publicly available information.

---

173 See Justices Clarify Scope of Anti-Hacking Law, supra note 20 (explaining what the Van Buren holding means for businesses). “This decision narrows the reach of the CFAA for such claims because misuse must have an element of exceeding access.” *Id.* See also Berris, supra note 12 (highlighting the vague meanings of authority in the CFAA). The concepts of “exceeds authorized access” and “without authorization” are undefined. *Id.* Authority to use a computer could be granted in many different ways; for example, “an employer who lets an employee use a work computer for business purposes or a website that allows users to access its servers for some function[,]” but the scope of authority is tricky to define and depends on the context of each individual situation. *Id.*

174 See Crook, supra note 119 (explaining the narrow interpretation of “without authorization”). “The Ninth Circuit held that ‘the concept of “without authorization” does not apply to public websites. … In the Ninth Circuit’s view, scraping public information from publicly accessible websites falls into the first category-and it is not an issue of authorization at all.” *Id.* “The Ninth Circuit’s interpretation is a positive development for those employing data scraping, but bad news for companies that seek to assert control over data posted on otherwise public websites.” *Id.* See Mackey & Opshal, supra note 90 (stating how the Van Buren decision is a victory for internet users). Van Buren’s holding “affirmed that online services cannot use the CFAA’s criminal provisions to enforce limitations on how or why you use their service, including for purposes such as collecting evidence of discrimination or identifying security vulnerabilities.” *Id.* “The Van Buren decision is especially good news for security researchers, whose work discovering security vulnerabilities is vital to the public interest but often requires accessing computers in ways that contravene terms of service.” *Id.*

175 See Crook, supra note 119 (outlining the court weighed business interests heavier than consumer privacy). “[T]he Ninth Circuit weighed hiQ’s business interest more heavily than LinkedIn’s stated interest in the privacy of its users, because the data at issue was comprised of profiles that these users chose to make public, and which LinkedIn itself disclaimed any responsibility for or ownership of.” *Id.*

176 See McRory, supra note 113 (suggesting there is a difference between public profile data and profile data that requires a log-in).
Though the application of the CFAA may vary based on the network involved, Congress should provide baseline distinctions on the level of interaction that a data scraper would need to take to meet the threshold of CFAA violation.\textsuperscript{177} “Without authorization” liability should include scenarios where the data scraper intentionally or knowingly retrieves non-public information through manipulating access.\textsuperscript{178} For example, there is a great difference between a users’ account that is entirely accessible from a simple Google search and a users’ account that is only accessible after passing through a log-in gate and becoming their friend.\textsuperscript{179} When the data scraper takes actions to retrieve personal information that is beyond the ease of immediate accessibility, their actions should fall within the “without authorization” zone.\textsuperscript{180} Furthermore, Congress should specify the point at which a user’s private data can be categorized as publicly available data.\textsuperscript{181} It is simple to say that if the data is accessible by

\textsuperscript{177} See Hasson, supra note 122 (suggesting that web scraping remains a tricky legal issue). An increasing number of organizations are adopting preventative measures to protect their data, while trying to maintain steady traffic flow to their website. \textit{Id.} See also McRory, supra note 113 (articulating the complex policy considerations). Due to the complexities and variations of website authentication, there is no one-size-fits-all approach. \textit{Id.} “In assessing how CFAA clarification can best meet policy goals, Congress must apply the CFAA differently for each social media platform.” \textit{Id.}

\textsuperscript{178} See 9-48.000 – COMPUTER FRAUD AND ABUSE ACT, supra note 14 (outlining the charging policy for access without authorization). “The Department will not charge defendants for accessing ‘without authorization’ . . . unless when, at the time of the defendant’s conduct, (1) the defendant was not authorized to access the protected computer under any circumstances . . . (2) the defendant knew of the facts that made the defendant’s access without authorization”. \textit{Id.}

\textsuperscript{179} See McRory, supra note 113 (arguing about the difference in public data). For profile data that can be viewed without a log-in, website owners would have no authority under the CFAA to ban bot scraping. \textit{Id.} For profile data that can only be viewed after passing through a log-in threshold, the issue of user privacy becomes the main concern. \textit{Id.}

\textsuperscript{180} See McRory, supra note 113 (describing how Congress could amend the CFAA to protect user privacy). Website owners should only have authority under the CFAA to ban data scrapers if the data they are collecting is not accessible after easily passing through a log-in gate. \textit{Id.} “If data collection involves any further interaction with users, then the social media network would have the authority under the CFAA to ban such activities.” \textit{Id.} This amendment would avoid unpredictable litigation outcomes and allow for greater user privacy. \textit{Id.}

\textsuperscript{181} See \textit{id.} (explaining that user privacy becomes a concern if the information can only be retrieved after passing through a log-in threshold, like on a Facebook account). See also hiQ Labs, Inc., 31 F.4th 2 at 1201 (rationalizing that “when a computer network generally permits public access to its data, a user’s accessing that
anyone with an internet connection, then it is publicly available; however, if the data scraper has to create an account with the respective network in order to view the data, is it still publicly available? Based on the Ninth Circuit’s holding in hiQ, creating an account to view users’ information does not likely equate to the “barrier of entry” trigger for CFAA violation; however, it is irrational to expect that every user’s account will be protected by a unique password gate, which would trigger a CFAA violation.183 Scraping plain view public information may not be enough for an entity to bring forward a CFAA action, but if the information requires additional actions, like an account to be created, a connection to be made, or a follower request to be sent, maybe that information is not so public after all.184

V. Conclusion

Our consumer data is everywhere. It is being bought and sold, scraped and tracked. It is being used to personalize products and increase bottom lines. As the legality of data scraping derives from the CFAA, Congress needs to be cautious of the implications the narrow interpretation of the statute can have. The CFAA should be amended to (1) penalize scrapers, beyond a barrier of entry, for any sort of manipulation of access and (2) clarify the threshold at which a publicly available data will not constitute access without authorization under the CFAA.

182 See Bajowala, supra note 110 (describing the CFAA as an anti-intrusion statute rather than a misappropriation statute). “[T]o violate the CFAA’s ‘without authorization’ provision, ‘an authentication requirement, such as a password gate, is needed to create the necessary barrier that divides open spaces from closed spaces on the Web.’” Id. See also McRory, supra note 113 (concluding that information accessible by anyone with a computer, “whether fully public or viewable after easily creating an account on the site, should be deemed public information that private social media companies have no authority to prevent collection of under the CFAA.”).

183 See Crook, supra note 119 (indicating that the concept of “without authorization” doesn’t apply to public websites). See also Bellia, supra note 89, at 1457 (referencing that authorization depends on whether a code-based system has been breached). See also hiQ Labs, Inc., 31 F.4th 2 at 1180 (holding Van Buren does not apply to public websites). Public websites lack limitations to access; anyone with an internet connection can access a public website. Id. There are no gates or barriers on a computer hosting publicly available webpages in the first place. Id.

184 See McRory, supra note 113 (reasoning that Congress should adjust the CFAA). Congress could provide CFAA clarifications by allowing public social media profile data without a log-in to not trigger CFAA liability and to narrow the scope of authority to scrape data that is only accessible after passing through a log-in threshold. Id.
user’s private information becomes public. The value of consumer data is only amplifying, yet there is not a single national regulation to protect it. A federal United States privacy law, similar to the European Union’s GDPR, is critical to ensure that companies are upholding adequate standards to safeguard consumers’ personal information. Ultimately, consumers deserve increased transparency from businesses, protection from the federal government, and the opportunity to decide how their personal data is used.