

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF COLUMBIA

CHRISTIAN W. SANDVIG
2117 Washtenaw Avenue
Ann Arbor, MI 48104,

KYRATSO KARAHALIOS
1109 S. Douglas Avenue
Urbana, IL 61801,

ALAN MISLOVE
5 Grayfield Avenue
West Roxbury, MA 02132,

CHRISTOPHER WILSON
46 Symmes Street, No. 3
Roslindale, MA 02131,

FIRST LOOK MEDIA WORKS, INC.
114 Fifth Avenue, 18th Floor
New York, NY 10011,

Plaintiffs,

-v.-

LORETTA LYNCH, in her official capacity as
Attorney General of the United States
950 Pennsylvania Avenue, NW
Washington, DC 20530,

Defendant.

Case No.

COMPLAINT
FOR DECLARATORY AND
INJUNCTIVE RELIEF

INTRODUCTION

1. This lawsuit challenges the constitutionality of a provision of the Computer Fraud and Abuse Act (“CFAA”), 18 U.S.C. § 1030 *et seq.*, a federal statute that prohibits and chills academics, researchers, and journalists from testing for discrimination on the internet. This chill arises because the CFAA makes it a crime to visit or access a website in a manner that violates that website’s terms of service, while

robust audit testing and investigations to uncover online discrimination require violating common website terms of service.

2. Without online audit testing, policymakers and the American public will have no way to ensure that the civil rights laws continue to protect individuals from discrimination in the twenty-first century.

3. In the offline world, audit testing has long been recognized as a crucial way to uncover racial discrimination in housing and employment and to vindicate the civil rights laws, in particular the Fair Housing Act (“FHA”) and Title VII’s prohibition on discrimination in employment. This testing involves pairing individuals of different races to pose as home- or job-seekers to determine whether they are treated differently. The law has long protected such socially useful misrepresentation in the offline world. In the online world, however, conducting the same kind of audit testing generally violates websites’ terms of service, which often prohibit providing false information, creating multiple user profiles, or using automated methods of recording the information displayed for different users.

4. The CFAA creates liability when an individual, in accessing a protected computer, does so in a manner that “exceeds authorized access.” 18 U.S.C. § 1030(a)(2)(C) (the “Challenged Provision”). Courts and federal prosecutors have interpreted the prohibition on “exceed[ing] authorized access” to make it a crime to visit a website in a manner that violates the terms of service or terms of use (hereinafter “terms of service” or “ToS”) established by that website. The Challenged Provision thereby delegates power to companies that operate online to define the scope of criminal law through their own terms of service. As a result, individuals and organizations risk

prosecution for conducting research into online discrimination where ToS prohibit their research techniques. They face prosecution even where, as in the case of Plaintiffs' activities, their research will not cause material harm to the target websites' operations and where they have no intent to commit fraud or to access any data or information that is not made available to the public.

5. The CFAA's prohibition on conducting robust research into online discrimination is of real concern given growing indications that proprietary algorithms are causing websites to discriminate among users, including on the basis of race, gender, and other characteristics protected from discrimination under the civil rights laws. Transactions involving the core social goods covered by federal and state civil rights laws—e.g., housing, credit, and employment—are increasingly taking place online. Simultaneously, actions on the internet are losing much of their anonymity, as “cookies” and other tracking technologies allow websites to access all kinds of information about visitors, including information that may reveal race, gender, age, and sexual orientation.

6. Companies that operate commercial websites have access to massive amounts of data about internet users and can employ sophisticated computer algorithms to analyze that data. Such “big data” analytics are used by many websites, and usage is constantly increasing and expanding. Big data enables behavioral targeting, meaning that websites can steer individuals toward different homes or credit offers or jobs—including based on their membership in a class protected by civil rights laws. Behavioral targeting opens up vast potential for discrimination against marginalized communities, including

people of color and other members of protected classes. The potential scope of this problem has been repeatedly acknowledged by the federal government.¹

7. The Plaintiffs in this case, academics and a media organization, wish to conduct audit testing or related investigative work to determine whether online websites—including those that advertise or provide a means by which individuals can apply for housing and employment—are treating users differently based on their membership in a protected class, but they are limited by the ToS of target websites. Some of the Plaintiffs have already engaged in such research and testing activities and must now fear prosecution under the Challenged Provision.

8. The Plaintiffs' research and testing activities, which include posing as online users of different races and recording the information they receive, constitute speech and expressive activity that is protected by the First Amendment, and that is prohibited by the Challenged Provision. The overbroad and indeterminate nature of the Challenged Provision prohibits and chills a range of speech and expressive activity protected by the First Amendment, because it prevents Plaintiffs and other individuals from conducting robust research on issues of public concern when websites choose to proscribe such activity.

¹ See Executive Office of the President, *Big Data: Seizing Opportunities, Preserving Values* 51-53 (May 2014), https://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf; Federal Trade Commission, *Big Data: A Tool for Inclusion or Exclusion?* (Jan. 2016), <https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf> (hereinafter "FTC Report on Big Data"); Executive Office of the President, *Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights* (May 2016), https://www.whitehouse.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf.

9. Plaintiffs therefore bring this action to enjoin the enforcement of the Challenged Provision, on its face and as applied to them, as violating the First Amendment and the Due Process Clause of the Fifth Amendment to the U.S. Constitution.

Jurisdiction and Venue

10. This action arises under the U.S. Constitution, including the First Amendment and the Due Process Clause of the Fifth Amendment.

11. This Court has jurisdiction over the subject matter of this action pursuant to 28 U.S.C. § 1331. This Court may award Plaintiffs declaratory and injunctive relief pursuant to the Declaratory Judgment Act, 28 U.S.C. §§ 2201-02, and this Court's inherent equitable jurisdiction.

12. Venue is proper in the U.S. District Court for the District of Columbia pursuant to 28 U.S.C. § 1391(b). Defendant, who is sued in her official capacity, resides in this judicial district. This action challenges the constitutionality of a statute that applies in this judicial district.

Parties

13. Plaintiff Christian W. Sandvig is an Associate Professor at the University of Michigan. He resides in Ann Arbor, Michigan.

14. Plaintiff Kyratso "Karrie" Karahalios is an Associate Professor at the University of Illinois. She resides in Urbana, Illinois.

15. Plaintiffs Sandvig and Karahalios are conducting a study to determine whether the computer programs that determine what housing to display on real estate websites are discriminating against users by race or other factors.

16. Plaintiff Alan Mislove is an Associate Professor at Northeastern University. He resides in West Roxbury, Massachusetts.

17. Plaintiff Christopher “Christo” Wilson is an Assistant Professor at Northeastern University. He resides in Roslindale, Massachusetts.

18. Plaintiffs Wilson and Mislove are conducting a study to test whether the ranking algorithms on major online hiring websites produce discriminatory outputs by systematically ranking specific classes of people (e.g., people of color or women) below others.

19. Plaintiff First Look Media Works, Inc. (“Media Works”) is the non-profit journalism arm of First Look Media, which has its principal place of business in New York, New York. First Look Media is a new-model media company devoted to supporting independent voices across all platforms. Media Works, a federally-recognized 501(c)(3) tax-exempt organization, publishes The Intercept, an online news and journalism platform. Its sister company, First Look Productions, Inc., produces and finances content for all screens and platforms including feature films, television, digital series, and podcasts.

20. Plaintiff Media Works and its journalists wish to engage in robust investigations of online companies and websites. They wish to investigate websites’ business practices and outcomes, including any discriminatory effects of websites’ use of big data and algorithms.

21. Defendant Loretta Lynch is the Attorney General of the United States and is sued in her official capacity. The Attorney General oversees the enforcement of federal

criminal statutes. As the head of the Department of Justice, she supervises its officers and employees, including the United States Attorneys.

The Computer Fraud and Abuse Act

22. The Computer Fraud and Abuse Act, 18 U.S.C. § 1030 *et seq.*, prohibits unauthorized access to “protected computer[s]” under certain circumstances.

23. The term “protected computer” includes a computer “which is used in or affecting interstate or foreign commerce or communication, including a computer located outside the United States that is used in a manner that affects interstate or foreign commerce or communication of the United States.” 18 U.S.C. § 1030(e)(2)(B).

24. A protected computer includes any website that is accessible on the internet. *See, e.g., United States v. Trotter*, 478 F.3d 918, 921 (8th Cir. 2007).

25. 18 U.S.C. § 1030(a)(2)(C) (the “Challenged Provision”) provides that:

Whoever . . . intentionally accesses a computer without authorization or exceeds authorized access, and thereby obtains . . . information from any protected computer . . . shall be punished as provided in subsection (c) of this section.

26. A first violation of the Challenged Provision carries a one-year maximum prison sentence and a fine. 18 U.S.C. § 1030(c)(2)(A). A second or subsequent violation carries a prison sentence of up to ten years and a fine. *Id.* § 1030(c)(2)(C).

27. The Challenged Provision contains no requirement of intent to cause harm, or of actual harm stemming from the prohibited conduct, before imposing criminal penalties.

28. While “without authorization” is not defined by the statute, “exceeds authorized access” means “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.” 18 U.S.C. § 1030(e)(6).

29. The “exceeds authorized access” language has been repeatedly interpreted by courts and the federal government to prohibit accessing a publicly-available website in a manner that violates that website’s terms of service.

30. The U.S. Department of Justice’s manual for CFAA prosecutions notes, in explaining the definition of the phrase “exceeds authorized access,” that it is “relatively easy to prove that a defendant had only limited authority to access a computer in cases where the defendant’s access was limited by restrictions that were memorialized in writing, such as terms of service [or] a website notice . . .” Office of Legal Education, Executive Office for United States Attorneys, Department of Justice, Prosecuting Computer Crimes, <http://www.justice.gov/criminal/cybercrime/docs/ccmanual.pdf>. It provides citations to caselaw for the proposition that violating such restrictions can suffice to prove the “exceeds authorized access” element of the Challenged Provision. *Id.* at 8–9.

31. The Department of Justice has brought at least two prosecutions alleging violations of 18 U.S.C. § 1030(a)(2) based on accessing a website in a manner that violates that website’s ToS. *See United States v. Drew*, 259 F.R.D. 449 (C.D. Cal. 2009); *United States v. Lawson*, No. CRIM. 10-114 KSH, 2010 WL 9552416 (D.N.J. Oct. 12, 2010).

32. The CFAA also provides for civil liability where a person “suffers damage or loss by reason of a violation” of its provisions. 18 U.S.C. § 1030(g). Courts adjudicating such civil actions have also interpreted “exceeds authorized access” to

encompass accessing information in violation of a website's terms of service. *See EF Cultural Travel BV v. Zefer Corp.*, 318 F.3d 58, 62 (1st Cir. 2003); *CollegeSource, Inc. v. AcademyOne, Inc.*, 597 Fed. App'x 116, 129–30 (3d Cir. 2015).

33. Criminal liability under the CFAA extends to “any individual, firm, corporation, educational institution, financial institution, governmental entity, or legal or other entity” that violates its provisions, 18 U.S.C. § 1030(e)(12), and to any of these for “conspir[ing] to commit” an offense, *id.* § 1030(b).

34. In addition to prohibiting the research and investigations that Plaintiffs wish to conduct, the Challenged Provision prohibits actions in furtherance of a plan to conduct such research and investigations. 18 U.S.C. § 1030(b).

35. Plaintiffs have an objectively reasonable belief that conducting the research and investigations they have designed to uncover discrimination online would subject them to criminal liability. They also have an objectively reasonable fear of criminal prosecution under the Challenged Provision.

Audit Testing and the Fair Housing Act

36. For more than three decades, testing has been central to enforcement of the Fair Housing Act (“FHA”), 42 U.S.C. § 3601 *et seq.* Testing has also played an important role in the enforcement of Title VII, 42 U.S.C. § 2000e *et seq.*, which prohibits discrimination in employment.

37. The Fair Housing Act has as its goal “to provide, within constitutional limitations, for fair housing throughout the United States.” 42 U.S.C. § 3601. To that end, the FHA prohibits discrimination in “the sale or rental of . . . a dwelling to any person because of race, color, religion, sex, familial status, or national origin.” 42 U.S.C.

§ 3604(a). It also prohibits actions that “otherwise make unavailable or deny” a dwelling to a person on those bases. *Id.*

38. The FHA further prohibits discrimination “in the terms, conditions, or privileges of sale or rental of a dwelling” on a prohibited basis, 42 U.S.C. § 3604(b), and the making of representations on a prohibited basis “that any dwelling is not available for inspection, sale, or rental when such dwelling is in fact so available,” *id.* § 3604(d).

39. The FHA also makes it illegal “[t]o make, print, or publish, or cause to be made, printed, or published any notice, statement, or advertisement, with respect to the sale or rental of a dwelling that indicates any preference, limitation, or discrimination based on” membership in a protected class. 42 U.S.C. § 3604(c).

40. The FHA prohibits both intentional discrimination and practices that, while facially neutral, disproportionately harm members of a protected class without sufficient justification. *See* 24 C.F.R. § 100.500 (practice has a prohibited discriminatory effect under the FHA “where it actually or predictably results in a disparate impact on a group of persons . . . because of race, color, religion, sex, handicap, familial status, or national origin,” and it is either not “necessary to achieve one or more substantial, legitimate, nondiscriminatory interests,” or any such interests “could be served by another practice that has a less discriminatory effect”).

41. Since the FHA’s passage, explicit statements of racial discrimination by housing providers and their agents have become much rarer, but discriminatory treatment and steering persist. Because it is nearly impossible for an individual to determine that she has been a victim of this more subtle discrimination without knowing about the

experiences of other prospective renters or buyers, paired testing has become the standard procedure for determining whether a housing provider is discriminating.

42. In a paired test, two people, one of whom is a member of a protected class and one of whom is not (e.g., a white tester and a Black tester) pose as equally qualified homeseekers and make the same inquiry about available homes. Multiple pairs may be sent to test the same housing provider or real estate agency.

43. Since the 1970s, the U.S. Department of Housing and Urban Development (“HUD”) has conducted a nationwide, comprehensive study of racial and ethnic discrimination in housing approximately once per decade. The most recent such study, published in 2013, applied paired-testing methodology in twenty-eight metropolitan areas and found that Black, Latino, and Asian testers were told about and shown fewer homes than white testers. U.S. Dep’t of Housing and Urban Development, Office of Policy Development and Research, *Housing Discrimination Against Racial and Ethnic Minorities 2012* xi, http://www.huduser.gov/portal/Publications/pdf/HUD-514_HDS2012.pdf.

44. The Supreme Court recognized that fair housing testers have standing to sue for FHA violations in *Havens Realty Corp v. Coleman*, 455 U.S. 363, 373 (1982). Courts regularly acknowledge the importance of testing to achieving the FHA’s aims. *See, e.g., Smith v. Pac. Properties & Dev. Corp.*, 358 F.3d 1097, 1102 (9th Cir. 2004); *Richardson v. Howard*, 712 F.2d 319, 321 (7th Cir. 1983).

45. Five years after *Havens*, Congress and the President affirmed and codified the importance of testing when Congress passed and the President signed the Housing and Community Development Act of 1987 (“HCDA”). Pub. L No. 100–242, 101 Stat

1815. The HCDA created the Fair Housing Initiatives Program, through which the Department of Housing and Urban Development funds private nonprofit fair housing enforcement organizations to enforce the FHA, including specifically “testing and other investigative activities” and “special projects, including the development of prototypes to respond to new or sophisticated forms of discrimination against persons protected” by the FHA. 42 U.S.C. §§ 3616a(b)(1), (b)(2)(A), (C).

Testing and Title VII

46. Title VII of the Civil Rights Act of 1964, 42 U.S.C. § 2000e *et seq.*, makes it illegal for an employer or an employment agency to engage in a number of prohibited employment practices because of the “race, color, religion, sex, or national origin” of an employee or prospective employee. 42 U.S.C. § 2000e-2(a)-(b).

47. Prohibited employer practices include refusing to hire, discharging, or applying different terms and conditions of employment to an individual because of a protected characteristic, and segregating individuals on those grounds. 42 U.S.C. § 2000e-2(a).

48. Title VII prohibits both intentional discrimination and any employment practice that causes a disparate impact on a prohibited basis if the practice is not “job related for the position in question and consistent with business necessity” or if there exists an “alternative employment practice” that could meet the employer or employment agency’s needs without causing the disparate impact. 42 U.S.C. § 2000e-2(k)(1).

49. Under Title VII, an employment agency is an entity “regularly undertaking with or without compensation to procure employees for an employer or to procure for employees opportunities to work for an employer and includes an agent of

such a person.” 42 U.S.C. § 2000e(c). Employment agencies may not “fail or refuse to refer for employment, or otherwise [] discriminate against” individuals on a protected basis. 42 U.S.C. § 2000e-2(b).

50. Paired testing for employment discrimination can be conducted in the form of correspondence tests or audit studies. In a correspondence test, auditors submit two job applications for fictional applicants that vary only with respect to racial or gender signifiers or other protected characteristics. In an in-person audit study, pairs of real testers apply for jobs, presenting equal credentials. *See* Devah Pager & Bruce Western, *Identifying Discrimination at Work: The Use of Field Experiments*, 68 J. of Social Issues 221, 223 (2012).

51. Recent paired tests of employment discrimination have consistently found white testers to receive approximately twice as many callbacks or job offers as Black testers. *Id.* at 226.

52. Courts have recognized the role of paired testing in the enforcement of Title VII. *Kyles v. J.K. Guardian Sec. Servs., Inc.*, 222 F.3d 289, 292 (7th Cir. 2000) (finding that recognizing testers’ standing in Title VII context “is consistent with the statute’s purpose”); *Fair Employment Council of Greater Washington, Inc. v. BMC Marketing Corp.*, 28 F.3d 1268, 1277–78 (D.C. Cir. 1994) (finding organization alleged a cause of action under Title VII against an employer based in part on evidence obtained by testers).

53. For more than two decades, the Equal Employment Opportunity Commission (“EEOC”) has also determined, based on caselaw and statutory construction, that testers have standing to bring claims of employment discrimination. EEOC Notice,

No. 915.002 (May 22, 1996), <http://www.eeoc.gov/policy/docs/testers.html>; EEOC Policy Guidance No. 915.062 (Nov. 20, 1990).

54. The above-described federal programs, under the auspices of federal civil rights statutes, as well as court cases upholding testers' standing and affirming the importance of testing, demonstrate the executive, congressional, and judicial understanding that such testing and investigations are socially valuable and, indeed, necessary.

The Need for Online Discrimination Testing and Investigation

55. In recent years, real estate, finance, and employment transactions have increasingly been initiated on the internet, and the trend will continue.

56. Simultaneously, the rise of "big data" has allowed for new forms of targeted marketing. Data brokers compile consumers' information from public records, social media sites, online tracking, and retail loyalty card programs and sell this information for marketing purposes.

57. Data brokers also place individual consumers into models that include inferences about them, including racial and ethnic inferences. Some of these models "primarily focus on minority communities with lower incomes, such as 'Urban Scramble' and 'Mobile Mixers' . . . which include a high concentration of Latino and African-American consumers with low incomes." Federal Trade Commission, *Data Brokers: A Call for Transparency and Accountability* 20 (May 2014), <http://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf>. Some segments are explicitly race-based, such as "African-American Professional" or

“Native American Lifestyle.” *Id.* at 21 and Appendix B-5. Other factors considered by data brokers are less explicit, but serve as proxies for race—such as “purchase behavior data” sorted by consumers interested in “Kwanzaa/African-Americana Gifts.” *Id.* at Appendix B-6. Data brokers also offer the ability to “append” additional information about consumers for retailers and other clients including race, age, gender, religion, and ethnicity. *Id.* at 24.

58. These profiles can follow individuals online, enabling websites and advertisers to display content targeted at, for example, African-American visitors or women.²

59. Tracking technologies, which allow websites and advertisers to compile records of individuals’ browsing histories, also allow for targeting. A “cookie,” for example, is a small piece of data sent from a web server to a user’s browser, stored there, and sent back from the browser on future requests to the same web server. Websites and advertisers can insert tracking cookies, thereby enabling them to see and analyze which websites a user has visited. Individuals who use websites that allow them to create accounts may have their browsing, purchasing, or social media history tracked and linked to their accounts. Such methods of tracking individuals allow companies to target marketing materials to them in the form of advertisements on websites that they visit, or

² See Tim McGuire, *et al.*, “Why Big Data is the New Competitive Advantage,” *IvyBusinessJournal* (Jul./Aug. 2012), <http://iveybusinessjournal.com/publication/why-big-data-is-the-new-competitive-advantage/>; U.S. Senate Committee on Commerce, Science, and Transportation, *A Review of the Data Broker Industry: Collection, Use, and Sale of Consumer Data for Marketing Purposes* (Dec. 18, 2013), http://www.commerce.senate.gov/public/_cache/files/0d2b3642-6221-4888-a631-08f2f255b577/AE5D72CBE7F44F5BFC846BECE22C875B.12.18.13-senate-commerce-committee-report-on-data-broker-industry.pdf; FTC Report on Big Data, *supra* note 1.

through selective display of opportunities on housing and employment websites, for example.

60. These tracking technologies make it possible for a website or advertiser to decide to show particular content to, for example, users who have visited the Black Entertainment Television (“BET”) website, or who have recently purchased feminine hygiene products, or who have clicked on an article about LGBT rights.

61. Given the long and deep history of racial discrimination in housing and employment, and the contemporary persistence of such discrimination, Plaintiffs and the public have reason to wonder whether this new technology is being harnessed for discriminatory purposes.

62. Moreover, when algorithms automate decisions, there is a very real risk that those decisions will unintentionally have a prohibited discriminatory effect on members of a protected class.

63. Scholars have identified various ways in which algorithms encode discrimination. *See., e.g.,* Solon Barocas & Andrew D. Selbst, *Big Data’s Disparate Impact*, 104 Calif. L. Rev. (forthcoming 2016). Algorithms seek to discern correlations in existing data sets in order to predict which factors correlate with desired outcomes. But the use of such algorithms could result in disparate outcomes for members of protected classes.

64. For example, if an existing data set concerning past hiring decisions reflects past discrimination, a hiring algorithm may avoid Latinos because Latinos were historically less likely to be hired. Similarly, if a data set reflects that people who live in certain zip codes are likely to have lower credit scores, a creditworthiness algorithm may

tag all people in those zip codes as less creditworthy, disproportionately affecting people of color who tend to live in poorer neighborhoods regardless of socioeconomic status. A patent has already been granted for a method enabling lenders to make credit decisions based on the credit ratings of members of an individual's social network. Such a patent has the potential to create a disparate impact in lending based on race.

65. When groups of people are systematically underrepresented in a data set, perhaps due to differential rates of internet access or use, error rates for those groups are likely to be higher, potentially causing additional discriminatory effects.

66. Additionally, a "machine learning" technique allows for modifying the analysis and outputs on the basis of the "training data," which may include outside data sources, individual user interactions with the website, or feedback from vendors or other corporate partners. The training data may itself change over time. In such cases, an algorithm may initially produce outcomes that are not discriminatory, but, over time, behavior on the part of users, vendors, or other data suppliers can "teach" the algorithm to discriminate in ways that harm members of protected classes.

67. Accordingly, although many advocates seek various forms of increased algorithmic transparency, the best way to determine whether members of protected classes are experiencing discrimination in transactions covered by civil rights laws is via outcomes-based audit testing, which enables researchers to discover how websites appear to different users. Without such outcomes-based audit testing of certain online websites there may be no way to determine whether discrimination is occurring.

68. These illustrative examples of the ways in which the use of algorithms and big data could lead to discrimination against members of protected classes are not exhaustive.

69. Given the risks of both intentionally and unintentionally discriminatory outcomes, online testing with the same goals as the testing that has long been conducted offline is necessary to enforce the civil rights statutes and to ensure their continuing viability. Additionally, new forms of journalistic investigations or academic research are crucial to discovering and documenting online business practices and outcomes that implicate civil rights and other matters of public concern.

The Impact of Website Terms of Service on Online Audit Testing

70. A common method of outcomes-based audit testing of algorithms involves using automated technology to access a website or other network service repeatedly, generally by creating false or artificial user profiles, in order to examine whether and how websites and the algorithms that govern them respond. But most websites' terms of service prohibit the use of automated technology and the creation of artificial user profiles, preventing researchers from auditing algorithms and publishing their findings. As a result, ToS effectively prohibit the use of the very research tools and methods that are necessary to determine whether discrimination is taking place.

71. For example, the ToS of Zillow.com, Trulia.com, Realtor.com, Redfin.com, Homes.com, Apartments.com, Curbed.com, ApartmentGuide.com, Hotpads.com, and ForRent.com— ten commonly-visited housing websites—all prohibit the automated recording of information from their sites (known as “scraping”). Homes.com even purports to prohibit manually copying any content or information

displayed on its website. Seven of those ten sites prohibit users from providing false information. Similarly, the ToS of LinkedIn.com, Indeed.com, Glassdoor.com, Monster.com, CareerBuilder.com, SimplyHired.com, SnagAJob.com, Beyond.com, Dice.com, and TheLadders.com—ten commonly-visited employment websites—all prohibit scraping and providing false information. Four of those employment websites prohibit the creation of more than one account.

72. Furthermore, a private website can set the conditions of when a visitor may speak about any information learned by visiting the website, including speech subsequent to visiting the site which is made in other forums. The private website can make it a condition of access to a website that a visitor never speak negatively about the website or company on another forum, for example, such as through a non-disparagement clause, but may allow positive speech about the website or company. Such non-disparagement clauses are often used in form contracts governing the sale of consumer goods or services, and they restrict consumers' ability to communicate regarding those goods or services. *See* Report of the U.S. Senate Committee on Commerce, Science, and Transportation on S. 2044: Consumer Review Freedom Act of 2015 (Dec. 8, 2015), <https://www.congress.gov/congressional-report/114th-congress/senate-report/175/1>.

73. Some websites have terms of service that require advance permission before using the site for research purposes. It is therefore far more likely that speakers who wish to portray a website in a positive light will receive authorization from the website owner to engage in research activities, including publication, than will those who wish to criticize the website. In such cases, speech critical of a website (e.g., speech

concerning discrimination in which a website may be engaged) is criminal, while speech supportive of that website is not, because the owner has authorized the latter.

74. It is virtually impossible for internet users to locate and read the content of the thousands of lengthy ToS to which they are subject. *See* Aleecia M. McDonald & Lorrie Faith Cranor, *The Cost of Reading Privacy Policies*, 4 I/S J.L. & Pol’y for Info. Soc’y 543, 558 (2009) (the average person visits 1,462 unique websites per year, each with their own terms of service); Casey Fiesler & Amy Bruckman, *Copyright Terms in Online Creative Communities* 2551, 2554 (ACM CHI Conference on Human Factors in Computing Systems, Working Paper, 2014) (reviewing ToS for 30 websites and finding that understanding them requires, on average, the reading level of a college sophomore and, collectively, would take nearly eight hours). Moreover, because websites frequently reserve the right to, and do, change their ToS without notifying users, even a user who did read and understand the complete ToS at the time she first used the website could be subjected to criminal liability for conduct that was not prohibited in the ToS at the time she read them. For example, of the twenty commonly used housing and employment websites listed above, 18 of those websites reserve the right to modify their ToS at any time.

Plaintiffs’ Research Plans

Christian W. Sandvig and Karrie Karahalios

75. Plaintiff Christian W. Sandvig is an Associate Professor at the University of Michigan in Ann Arbor, Michigan. He is appointed at the Department of Communication Studies within the College of Literature, Science, and the Arts and the School of Information.

76. Plaintiff Sandvig holds a Ph.D. in Communications from Stanford University.

77. Plaintiff Sandvig's research investigates communication and information technology infrastructure and public policy. Among other areas, he focuses on algorithmic discrimination in the online context.

78. Plaintiff Karrie Karahalios is an Associate Professor of Computer Science at the University of Illinois.

79. Plaintiff Karahalios holds a Ph.D. in Media Arts and Sciences from the Massachusetts Institute of Technology.

80. Plaintiff Karahalios' research focuses on social computing, social network analysis, social spaces, and smart infrastructure.

81. Plaintiffs Sandvig and Karahalios have frequently collaborated; they have jointly conducted multidisciplinary research studies that investigate the potential for harmful online discrimination by internet platforms. Both plaintiffs are affiliates of The Center for People & Infrastructures at the University of Illinois, a research center dedicated to this purpose.

82. Plaintiffs Sandvig and Karahalios are in the process of designing and conducting a study that would investigate whether the computer programs that determine what to display on real estate websites are discriminating against users by race or other factors.

83. Online residential real estate websites ("sites") maintain a database of available properties by purchasing property listings from multiple listing services; they also may allow landlords, brokerages and realtors to directly submit listings. These

“organic listings” are usually displayed in response to a search query (e.g., “Capitol Hill” or “1 bedroom”). The same listings are typically displayed for every visitor that makes the same query at the same time.

84. These sites make money by accepting advertising, much of which advertises available properties and related services (such as real estate agents and mortgages). Some advertising is managed directly by the sites themselves, but real estate sites also participate in online advertising exchanges and networks, which manage a large inventory of advertisements for a variety of products and decide which advertisements to display on a designated portion of a web page. When individuals visit real estate websites to search for housing, they are shown properties from all of the above sources, including both organic listings and advertisements.

85. Online advertising networks and exchanges show different advertisements to different people. It is thus much more likely that advertising networks and exchanges could unintentionally discriminate in a harmful way: they already profile users to determine what advertisements to show them.

86. To study this problem in the context of racial discrimination, Plaintiffs Sandvig and Karahalios will vary the race of multiple “visitors” to real estate sites and measure any corresponding differences in the properties they are shown, holding other potential differences between visitors constant.

87. Plaintiffs Sandvig and Karahalios will first determine how race correlates with behavior that could be detected by an advertising network’s profiling apparatus by consulting published research and marketing statistics, which indicate that certain Web browsing behaviors are very highly associated with particular races. They will then

identify sites that are very likely indicators of race and the subset of these sites that participate in the same advertising networks used by online residential real estate sites, such that the networks use visits to these sites to determine which advertisements to show a user who later visits a real estate site.

88. Plaintiffs Sandvig and Karahalios will write a computer program that will act as though it is a real person browsing the Web. This program is an automated program or agent browsing the Web, referred to as a “bot.” Each bot represents an individual person and is designed to interact with a website as a user might. It can visit websites, click links, fill out and submit forms, collect and store information from a web page, and do other things automatically, based on scripts written by Plaintiffs Sandvig and Karahalios.

89. The bot will be instructed to behave as a number of different users; each of these profiles is a “sock puppet.”

90. The bot will first visit an online residential real estate site and search for properties, recording the properties offered to the sock puppet via advertising. Both the organic listings displayed and properties offered in advertisements to the bot will constitute the baseline for the experiment.

91. Plaintiffs Sandvig and Karahalios will then instruct the bot to perform the exhibiting behaviors associated with a particular race, so that, for instance, one sock puppet would browse like a Black user, while another would browse like a white user. All the sock puppets will browse the Web for several weeks, periodically revisiting the initial real estate site to search for properties.

92. At each visit to the real estate site, Plaintiffs Sandvig and Karahalios will record the properties that were advertised to that sock puppet by scraping that data from the real estate site. They will scrape the organic listings and the Uniform Resource Locator (“URL”) of any advertisements. They will also record images of the advertisements shown to the sock puppets.

93. Finally, Plaintiffs Sandvig and Karahalios will compare the number and location of properties offered to different sock puppets, as well as the properties offered to the same sock puppet at different times. They seek to identify cases where the sock puppet behaved as though it were a person of a particular race and that behavior caused it to see a significantly different set of properties, whether in number or location.

94. A finding of automated discrimination by online residential real estate websites would produce important new scientific knowledge about the operation of computer systems, discrimination, and cumulative disadvantage.

95. Plaintiffs Sandvig and Karahalios are aware that this experimental design will violate websites’ terms of service. The use of bots is prohibited by many websites that the bot would visit in the course of building the racially-identifiable sock puppets. Scraping is prohibited by the terms of service of virtually all real estate websites. The particular real estate websites they will test prohibit scraping.

96. This experimental design will have no impact, or at most a minimal impact, on the target websites’ operations.

97. Plaintiffs Sandvig and Karahalios were among the authors of a 2014 paper concerning methods for auditing algorithmic discrimination, in which they expressed their concerns about liability under the CFAA. Christian Sandvig, Kevin Hamilton,

Karrie Karahalios, & Cedric Langbort, *Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms* 12–13, May 22, 2014.

98. Plaintiffs Sandvig and Karahalios are concerned that violating terms of service in the course of their work will subject them to criminal prosecution under the Challenged Provision. Plaintiffs Sandvig and Karahalios have already begun some of the activities that are part of their research plan described above, including the activities that require violating websites' terms of service. They plan to continue to engage in this research because they believe it to be socially valuable and important.

99. Plaintiffs Sandvig and Karahalios do not wish to be exposed to criminal prosecution as a result of conducting research into online discrimination.

Alan Mislove and Christopher Wilson

100. Plaintiff Alan Mislove is an Associate Professor of Computer Science at Northeastern University, in the College of Computer and Information Science.

101. He holds a Ph.D. in Computer Science from Rice University.

102. Plaintiff Mislove's research investigates systems, networking, network measurement, and security and privacy issues associated with online social networks. Among other areas, he focuses on auditing the algorithms of large-scale systems.

103. Plaintiff Christopher ("Christo") Wilson is an Assistant Professor of Computer Science at Northeastern University, in the College of Computer and Information Science.

104. He holds a Ph.D. in Computer Science from the University of California, Santa Barbara.

105. Plaintiff Wilson’s research focuses on auditing algorithms, security and privacy, and online social networks.

106. Plaintiffs Mislove and Wilson have frequently collaborated. They work together as part of the Algorithmic Auditing Research Group at Northeastern University and have co-authored several papers measuring personalization and discrimination online. They have used the knowledge gained from measurements of the internet to build systems that improve security, privacy, and transparency for internet users.

107. Plaintiffs Mislove and Wilson plan to conduct research into algorithmic discrimination in the employment context. They have designed and plan to conduct a study that would determine if the algorithms used by hiring websites produce results that discriminate against job seekers by race, gender, and other factors.

108. Job seekers create personal profiles on online hiring websites, upload their resumes, and apply for open positions. At the same time, companies and recruiters post open positions onto these sites, and use the sophisticated tools the websites provide to screen candidates.

109. Hiring websites provide recruiters with a search engine-like interface that allows recruiters to query, filter, and browse all of the job seekers on the website. Like any search engine, these recruiter tools use proprietary algorithms to rank job seekers by opaque measures of “relevance.” The order of the ranking may influence who gets offered employment and who is passed over.

110. Plaintiffs Mislove and Wilson seek to determine whether the ranking algorithms on major online hiring websites produce discriminatory outputs by

systematically ranking specific classes of people (e.g., people of color or women) below others. This could happen intentionally or inadvertently.

111. Their audit study will test the hypothesis that these hiring websites may produce discriminatory outputs by relying on data that includes real-world biases. For example, a hiring website could rank job candidates in search results in a racially disparate manner if the algorithm that determines which results are displayed take into account factors—gleaned from a user’s resume, browsing history, or social networking profiles—that correlate with race.

112. On each hiring website, Plaintiffs Mislove and Wilson will investigate whether there are correlations between the rank ordering of job candidates in search results and race, gender, or age. If they observe that candidates with specific attributes are consistently ranked lower, this may indicate that the algorithm is discriminatory.

113. To investigate this problem in the context of racial discrimination, they will employ a hybrid auditing methodology.

114. First, in the observational stage of the study, they will create baseline demographic data by “crawling” a large random sample of users on the target websites using a bot. A bot can, among other things, visit websites and click links automatically, based on scripts written by Plaintiffs Mislove and Wilson. The bot will allow them to gather information about the random sample of users.

115. Second, Plaintiffs Mislove and Wilson will create employer accounts and then systematically run queries for job seekers, recording the ranked lists of candidates returned by the search engine. They will vary the keywords used in searches (e.g., “programmer,” “software developer,” “software engineer”) as well as the search filters

(e.g., years of experience, previous employer). Once ranked lists of candidates are returned in response to search queries, they will “scrape” the website as a method of recording the lists of candidates. Scraping means that the webpages returned by the search engine will be stored to a hard drive, and relevant information (e.g., the ranked list of job candidates) will be automatically extracted from the stored pages by software.

116. In addition to recording the ranked lists of candidates returned by the search engines, Plaintiffs Mislove and Wilson will crawl each candidate’s personal profile to collect any available information such as age, location, education, and experience. They will then label the attributes of users based on their profile data. To obtain a label for the race of each candidate, they plan to have multiple people label each user’s photograph. They will then quantify the distribution of users by race.

117. Plaintiffs Mislove and Wilson will then build statistical models that attempt to explain the observed rank orderings of candidates in search results, examining the impact of demographic variables (including race) on rank in the search results.

118. Second, in the experimental stage of their study, Plaintiffs Mislove and Wilson will create profiles for fictitious job seekers, post fictitious job opportunities, and have the fictitious users apply for the fictitious jobs. The goal in this phase is to examine how the websites rank the fictitious candidates who apply for their fictitious jobs.

119. Plaintiffs Mislove and Wilson will create sock puppet job-seeker accounts with varying attributes (e.g., race, gender, age). These accounts will always include one uniform, globally unique attribute (e.g., attendance at a fictitious high school) so that they can search for the sock puppets as distinct from genuine jobseekers. These sock puppet accounts will then be used to search for fictitious jobs that they will post on the websites.

120. Plaintiffs Mislove and Wilson will use all available mechanisms to prevent real people from applying for their fictitious jobs, including giving the job an explicit title (e.g., “This is not a real job, do not apply”). Once the experiments are over, they will delete all of the fictitious accounts and jobs that were created.

121. Plaintiffs Mislove and Wilson will systematically conduct searches as the employers of the fictitious jobs they have posted (in other words, using the employer accounts they have created), using the same keywords that were tested earlier, with the addition of a filter corresponding to the uniform attribute (e.g., the fictitious high school). This will ensure that the search results contain only the sock puppets.

122. By comparing the sock puppets’ rankings to the baseline search result distributions (i.e., those they found in their observational study), Plaintiffs Mislove and Wilson will be able to examine how specific user attributes—in this case, race—impact search rank.

123. In addition to publishing their findings in academic papers, Plaintiffs Mislove and Wilson plan to bring the results of this research to the public. They plan to develop a tool that will analyze a person’s profile on major hiring sites and rank it compared to various sock puppets. The tool will rank users using the same features as the true algorithms. The tool will teach people about the algorithms underlying hiring sites.

124. Plaintiffs Mislove and Wilson are aware that this experimental design violates websites’ terms of service. Use of crawling and scraping is prohibited by many of the websites that they would crawl or scrape to develop baseline data or record results. The use of sock puppets is prohibited by the terms of service of all hiring websites, which prohibit users from creating profiles containing false information.

125. This experimental design will have a minimal impact, if any, on the target websites' operations.

126. Plaintiffs Mislove and Wilson are concerned that violating terms of service in the course of this work will subject them to criminal prosecution under the Challenged Provision. Plaintiffs Mislove and Wilson have already begun some of the activities that are part of their research plan described above, including those activities that require violating websites' terms of service.

127. Plaintiffs Mislove and Wilson plan to continue to engage in this research because they believe that it may have significant social value. First, individual algorithm audits may uncover harmful discriminatory practices that, once exposed, force the relevant parties to change their behavior. This may also deter other organizations from using similar algorithms. Second, the tools and data that Plaintiffs Mislove and Wilson create during this project will aid academics and regulators who wish to expand on their findings or conduct their own audits. Finally, by educating computer scientists and the general public, Plaintiffs Mislove and Wilson hope to inform an important societal debate about the role and norms of algorithms in daily life.

128. Plaintiffs Mislove and Wilson do not wish to be exposed to criminal prosecution as a result of conducting research into online discrimination.

Plaintiff First Look Media Works

129. Media Works conducts investigative journalism through The Intercept, a website that publishes long-form investigative articles based on its journalists' original reporting and research. Among the subject matters of interest to the journalists at The Intercept are criminal justice, corporate practices, national security, and technology. The

Intercept has published a series of articles (now collected in a book, *The Assassination Complex: Inside the Government's Secret Drone Warfare Program*) about the United States' use of targeted drone attacks, and a multi-part investigation of how the DuPont company harmed communities' water sources while manufacturing Teflon. Working with large data sets has been a key component of many of The Intercept's articles.

130. Plaintiff Media Works and its journalists wish to engage in robust investigations of online companies, websites, and platforms. They wish to investigate websites' business practices and outcomes, including any discriminatory effects of websites' use of big data and algorithms.

131. Plaintiff Media Works and its journalists wish to violate certain website terms of service in order to conduct their investigations, including by scraping data from websites that is available either to the general public or to individual website users.

132. Plaintiff Media Works does not wish to be exposed to criminal prosecution as a result of engaging in necessary journalistic activity in order to inform the public about online business practices.

Plaintiffs' Injuries

133. Plaintiffs Sandvig, Karahalios, Misllove, and Wilson have the goal of conducting research and testing that would determine whether housing or employment websites are discriminating based on race, gender, or membership in other protected classes. Plaintiff Media Works has the goal of engaging in investigative journalism to research websites' business practices and outcomes.

134. Plaintiffs wish to gather and analyze data that is made available by the targeted websites and report on their findings to the public.

135. The research, testing, and investigative methods they have designed would, if carried out, violate the Challenged Provision, because they all require violating the terms of service of the targeted website. The research, testing, and investigative methods Plaintiffs wish to conduct would not be criminal but for the Challenged Provision.

136. None of the Plaintiffs' activities are done with the intent to defraud or cause material harm to any targeted website's operations, but instead with the intent to determine whether targeted websites are engaging in discrimination. To the extent, if any, that Plaintiffs' activities might burden a website's operations, the burden would be de minimis. To the extent any reputational or similar harm arises from Plaintiffs' subsequent publication of truthful information about their research findings, including any findings of discrimination, the government does not have a legitimate interest in preventing such harm.

137. Plaintiffs are injured because they are placed in the position of either refraining from conducting their research, testing, or investigations—all of which constitute constitutionally-protected speech or expressive activity, or conduct necessarily antecedent to such speech or expressive activity—or of exposing themselves to the risk of prosecution under the Challenged Provision. Refraining from conducting their research, testing, or investigations constitutes self-censorship and a loss of Plaintiffs' First Amendment rights.

138. Plaintiffs Sandvig, Karahalios, Mislove, and Wilson have already begun some of the activities described in their research plans, which include those activities that

violate websites' terms of service, and they reasonably fear prosecution under the Challenged Provision.

139. The Challenged Provision chills the Plaintiffs and others who wish to conduct similar online research, testing, or investigations for the following reasons:

1) because they are placed in reasonable fear of being prosecuted for engaging in constitutionally-protected expressive activity to uncover and report on discrimination and related matters, or 2) because they must alter or modify their research and testing design in a manner that may be less methodologically rigorous to accommodate terms of service in a way that reduces or eliminates their risk of prosecution, or 3) because they must refrain from conducting research or testing that violates websites' terms of service to avoid the risk of prosecution.

The Challenged Provision Violates the First Amendment on Its Face and As Applied

140. The Challenged Provision is unconstitutionally overbroad.

141. The Challenged Provision impermissibly burdens speech about business practices and other activity on the internet because websites can determine what speech and expressive activity to prohibit, and these prohibitions become criminal violations of the Challenged Provision. In other words, a website can *explicitly* target speech or expressive activities. For example, if a website's terms of service provide that access by certain types of speakers (such as researchers) is unauthorized, or that engaging in certain speech (false or misleading speech, for example, or subsequent disparaging speech about the website) renders access unauthorized, then violations of those private terms of service become crimes through the phrase "exceeds authorized access" in the Challenged

Provision. Such speech or expressive activity thus becomes prohibited under pain of criminal sanctions simply because it occurred on the internet.

142. Because the Challenged Provision incorporates websites' terms of service into the federal criminal code, its applications are virtually infinite; any speech or expressive activity that the private operator of a website has prohibited as a condition of access to its website becomes a criminal violation, even where that prohibition covers speech subsequent to the visit and in a different forum. In a good number of cases, a website's ToS will prohibit speech that cannot constitutionally be prohibited. Accordingly, although the Challenged Provision may have legitimate applications, its unconstitutional applications are substantial in relation to its legitimate scope.

143. The Challenged Provision is also overbroad because it prohibits a wide variety of conduct that is commonplace on the internet, from overstating one's height to copying information from real estate or job listings to be shared with others.

144. The Challenged Provision is overbroad because it prohibits the Plaintiffs' activities. Plaintiffs wish to gather, disseminate, and publish information about discrimination, activities constituting speech under the First Amendment but prohibited by the Challenged Provision at a website's behest. In order to gather the necessary information, they wish to create artificial "tester" profiles, violating ToS prohibitions on populating accounts with false information.

145. Plaintiffs wish to engage in anonymous speech and misrepresentation for the purpose of testing for discrimination. In this context, anonymous speech and misrepresentation enjoy First Amendment protection. However, the Challenged Provision renders such anonymous speech and misrepresentation criminal, simply because the

tester is evaluating an online business that has terms of service prohibiting such activity, or because the online business does not wish to be the target of such testing.

146. Some of Plaintiffs' research, testing, or investigation entails automated recording or collection of publicly-available information from websites, prohibited as data "scraping" by terms of service, but protected by the First Amendment.

147. Plaintiffs also wish to use websites for research purposes and to have the option of subsequently publishing the findings of their research, even when website terms of service do not allow doing so. However, the Challenged Provision renders such activities criminal.

148. The Challenged Provision's broad delegation of criminal regulation to private parties also impairs the First Amendment rights of many other people.

149. Terms of service, including those on social media websites, often require that users provide their real names when creating accounts, as is the case with seven of the ten housing websites and all of the employment websites listed in paragraph 71 of this Complaint. By criminalizing any violation of these rules, the Challenged Provision chills a broad range of important expressive activity. Members of marginalized groups, or victims of abuse and harassment, may seek to operate pseudonymously online in order to protect themselves. For example, real name policies chill the speech of lesbian, gay, bisexual, or transgender individuals who wish to keep this aspect of their identities private or separate from their offline lives, and they chill the speech of victims of domestic violence who would speak online but for fear of response from abusers. Transgender individuals whose legal names do not reflect their gender identities may be deterred from speaking online under such policies. Critics of employers, governments, or

other powerful actors may desire the safeguard from retribution that pseudonymity provides. Artists, writers, and others engaged in creative expression often desire pseudonymity, and some forms of satire, such as fictional Twitter accounts, depend upon misrepresenting the user's identity. The Challenged Provision criminalizes violations of websites' real name policies, chilling this entire range of constitutionally protected activity.

150. The Challenged Provision also criminalizes violations of websites' requirements that users provide truthful information in other aspects of their profiles. Thus, it threatens prosecution for false speech about a dating website user's age, height, or weight, and for a false declaration of party affiliation by a commenter on a news website, to name just two examples.

151. Some websites' terms of service explicitly prohibit criticism of the website or company in any forum as a condition of use. Such a straightforward prohibition on speech is also rendered criminal by the Challenged Provision.

152. Many more unconstitutional applications of the Challenged Provision exist, including those stemming from terms of service prohibitions on recording publicly-available content on websites, or sharing or providing access to information available through use of personal accounts.

153. Because it incorporates websites' prohibitions on speech and expressive activity, including speech and expressive activity protected by the First Amendment, the Challenged Provision prevents or chills speakers from exercising their First Amendment rights online and is overbroad.

154. The Challenged Provision violates the First Amendment as applied to the Plaintiffs. In order to conduct their proposed research, testing, and investigations, Plaintiffs wish to engage in protected speech or expressive activity prohibited by terms of service, in turn rendered criminal by the Challenged Provision. The online research that Plaintiffs wish to conduct includes accessing websites using artificial tester profiles, in violation of terms of service that prohibit providing false information. But such conduct enjoys First Amendment protection.

155. The freedom to conduct academic research, and the freedom of the press, are of paramount public importance and entitled to full protection under the First Amendment. The Plaintiff researchers and journalists wish to study the subject of online discrimination using the methodologies of their professions, and they should not be restricted in using otherwise lawful tools and techniques simply because they pursue their research on the internet.

156. As applied to Plaintiffs, the Challenged Provision fails strict scrutiny. The government, far from having a compelling interest in preventing Plaintiffs' speech and expressive activity, has an interest in ensuring the enforcement of anti-discrimination laws online. The government's interest in preventing computer crime is more than adequately served by other provisions of the CFAA, which prohibit accessing protected computers when, *inter alia*, damage is caused or there is an intent to defraud. *See, e.g.*, 18 U.S.C. §§ 1030(a)(4), (a)(5)(B), (a)(5)(C).

157. Plaintiffs also wish to use automated methods of recording publicly-available data or data available to individual users from the audited websites that are

prohibited by terms of service, even though such recording constitutes protected First Amendment activity.

158. Plaintiffs wish to record algorithms' outputs using an automated "scraping" technique, which allows for rapid gathering of large amounts of data that would take far longer to gather manually. Automated scraping is generally barred by ToS. Similarly, ToS often restrict sharing information gleaned by an individual account holder, by, for example, prohibiting the sharing of passwords which would allow multiple people to view the information that the site displays to a particular user.

159. Recording and retaining publicly-available information from websites, like video or audio recording of public places, is expressive activity protected by the First Amendment. Furthermore, sharing information that a website makes available to a particular user is critical to comparing outputs based on membership in a protected class, and constitutes speech or recording protected by the First Amendment.

160. Plaintiffs wish to have the option of publishing the results of their research, including any findings of discrimination, even if a target website's ToS prohibit doing so. Such publication is protected by the First Amendment.

161. As applied to the Plaintiffs in conducting their proposed research plans and journalistic activities, the Challenged Provision violates the First Amendment.

The Challenged Provision Criminalizes Speech Necessary to Petition the Government

162. The Challenged Provision makes it a criminal violation to petition the government for a redress of grievances where a website's terms of service prohibit the speech necessary to engage in such petitioning.

163. For example, if a website conditions access based on a requirement that the user not make any subsequent negative, critical, or disparaging speech about that website, then the user cannot report discrimination by that website to the government.

164. Where visiting or using a website triggers a ToS restriction on subsequent speech to petition the government, then no petition for redress of grievances can be based on the type of research Plaintiffs wish to conduct. Without this type of research, it is impossible to determine whether housing and employment websites, and the businesses that post and advertise through them, are violating Fair Housing Act or Title VII rights, or are otherwise discriminating against members of groups protected by the civil rights laws. By preventing individuals from subsequently speaking about such research, the Challenged Provision precludes knowledge of the scope and extent of online discrimination.

165. This is especially true when algorithms are in the position to automate discrimination. Even if the private companies that operate housing- and employment-related websites could somehow be compelled to share the computer code underlying their algorithms, that code would not give a full picture of how the algorithm works in practice. Some additional factors that could influence outcomes include: unshared or dynamic datasets; interactions with outside vendors; and patterns of behavior that arise from interactions with users. Moreover, some modern algorithms (including many machine-learning algorithms) can be both dynamic and complex, such that they are simply not comprehensible to any human auditor at any particular point in time by looking at the code itself. Observations and analysis of algorithmic behavior are

necessary in these cases to understand the nature of the constructed systems, and such observations and analysis necessitate visiting a website.

166. In other words, online websites, by controlling their terms of service, can control whether or not potentially adverse information about their practices is reported to the government by any user who has ever visited.

167. The Challenged Provision bars engaging in legislative and administrative advocacy, or in litigation, alleging discrimination by a website where its ToS prohibit such advocacy or litigation. Thus individuals cannot identify for HUD, the EEOC, or other relevant agencies the particular discrimination problems that exist on a particular website and lobby those agencies for rules or other guidance that would ensure the robust enforcement of current law online. Should online audit testing reveal discrimination that falls outside the reach of the existing law, individuals could not lobby Congress for new protections specific to the online context.

168. Individuals could not access the courts to enforce Fair Housing Act and Title VII rights after visiting a website with ToS that prohibit doing so. Thus, any victim of online discrimination by a website, who must necessarily have visited or used that website, will be precluded from making a claim of discrimination and will be unable to pursue such a claim in court.

169. The Challenged Provision essentially delegates to potential defendants in such lawsuits the ability to prevent speech about relevant evidence. Because it is these potential defendants who draft terms of service, violations of which the Challenged Provision renders criminal, the recipe for avoiding Fair Housing Act and Title VII liability for algorithmic discrimination is straightforward: merely employ terms of service

that preclude subsequent speech about such discrimination, and it can continue unchecked.

The Challenged Provision is Vague, In Violation of the Due Process Clause

170. The Challenged Provision, which prohibits accessing a protected computer in a manner that “exceeds authorized access” is, on its face, void for vagueness.

171. The Challenged Provision fails to notify ordinary people of what conduct is criminal because the phrase “exceeds authorized access” does not provide sufficient notice that an individual must comply with a website’s written terms of service at all times. The plain meaning of the phrase “exceeds authorized access” does not clearly cover instances where a website places no barriers, such as technological or physical barriers, to access by individuals.

172. Because the Challenged Provision chills speech and expressive activity as described above, the Due Process Clause requires a heightened degree of statutory specificity. The vagueness of the Challenged Provision fails to give reasonable notice of what conduct is prohibited, invites arbitrary and discriminatory enforcement, and deters constitutionally-protected speech. It thus violates the Due Process Clause.

The Challenged Provision Represents an Unconstitutional Delegation of Authority to Private Parties

173. The Challenged Provision delegates to website owners the legislative power to determine which conduct is criminal.

174. The Challenged Provision makes it a federal crime to visit a website in a manner that “exceeds authorized access.” The private parties that draft terms of service determine the conditions under which access is authorized; as a result, they wield the power to define the conduct that violates the Challenged Provision, including conduct

that occurs subsequent to accessing a website or is unrelated to any legitimate access restriction.

175. The Challenged Provision does not merely provide for the enforcement of private contractual arrangements: It renders conduct a separate, federal crime if it violates a website's ToS.

176. The private processes through which terms of service are drafted and approved are closed and nontransparent, with no requirement for public comment or participation. Because terms of service can be and are constantly revised, members of the public lack even the most basic notice that revisions are in progress, and have no right to participate in defining what terms of service require.

177. The government retains no control over the lawmaking process because terms of service prohibitions, drafted by private parties without public input, effectively become criminal prohibitions backed by federal law. The Challenged Provision allows private parties unilaterally and undemocratically to define the conduct that constitutes a crime.

178. The Challenged Provision fails to notify ordinary people of what conduct is criminal because there is no requirement that ToS be drafted with the requisite clarity or precision required for defining conduct that is criminal.

179. For these reasons, the Challenged Provision's delegation of the legislative power to private parties completely removes the lawmaking function from the political process and from the mechanisms for democratic accountability, and is unconstitutional.

CLAIMS FOR RELIEF

First Cause of Action

FREE SPEECH

[U.S. Const., amend. 1 (Freedom of Speech and Freedom of Press Clauses)]

180. Plaintiffs re-allege and incorporate by reference all allegations set forth above.

181. The Free Speech and Free Press Clauses of the First Amendment to the U.S. Constitution provide: “Congress shall make no law . . . abridging the freedom of speech, or of the press.”

182. The Challenged Provision prevents speech and expressive activity necessary to inform and influence the decisions of the public and the government on online discrimination.

183. The Challenged Provision is unconstitutionally overbroad. By prohibiting access to websites that “exceeds authorized access,” the Challenged Provision incorporates the terms of service of each and every website into its text. It thereby creates virtually limitless restrictions on speech and expressive activity, including the speech and expressive activity that Plaintiffs here wish to engage in. The Challenged Provision is unconstitutionally overbroad on its face because its unconstitutional applications are substantial in relation to its legitimate applications.

184. As applied to the Plaintiffs, the Challenged Provision unconstitutionally restricts their protected speech, recording activities, and other protected expressive activities as described above. The Plaintiffs’ research plans and journalistic activities are

not done with the intent to cause harm to any target websites' operations, and any harm that may result is de minimis.

185. The Challenged Provision is not narrowly tailored to any legitimate, compelling, or overriding government interest. The government in fact has an interest in the completion of Plaintiffs' research, an interest expressed through the Fair Housing Act and Title VII.

186. The Challenged Provision violates the Free Speech and Free Press Clauses of the First Amendment.

Second Cause of Action

RIGHT TO PETITION

U.S. Const., amend. 1 (Petition Clause)

187. Plaintiffs re-allege and incorporate by reference all allegations set forth above.

188. The Petition Clause of the First Amendment of the U.S. Constitution provides that "Congress shall make no law . . . abridging . . . the right of the people . . . to petition the Government for a redress of grievances." U.S. Const., amend. I.

189. The Challenged Provision prohibits the speech necessary to communicate with HUD, the EEOC, and other federal and state government entities concerning the enforcement of the Fair Housing Act, Title VII, and other civil rights laws online. Such communications are protected by the Petition Clause.

190. The Challenged Provision prohibits the speech necessary to access the courts in order to enforce rights granted by the Fair Housing Act, Title VII, and other civil rights laws in the online context.

191. The Challenged Provision is not justified by a legitimate, compelling, or overriding government interest.

192. The Challenged Provision is not narrowly tailored to achieve any such legitimate, compelling, or overriding government interest.

193. The Challenged Provision violates the Petition Clause of the First Amendment.

Third Cause of Action

VOID FOR VAGUENESS

U.S. Const., amend. 5 (Due Process Clause)

194. Plaintiffs re-allege and incorporate by reference all allegations set forth above.

195. The Due Process Clause of the Fifth Amendment to the United States Constitution provides that “No person shall . . . be deprived of life, liberty, or property, without due process of law.”

196. The Challenged Provision is unconstitutionally vague, as it fails to define a criminal offense in a manner definite enough to notify an ordinary person what conduct is prohibited.

197. The vagueness of the Challenged Provision chills and deters speech and expressive activity protected by the First Amendment.

198. The Challenged Provision violates the Due Process Clause of the Fifth Amendment.

Fourth Cause of Action

UNCONSTITUTIONAL DELEGATION

U.S. Const., amend. 5 (Due Process Clause)

199. Plaintiffs re-allege and incorporate by reference all allegations set forth above.

200. The Due Process Clause of the Fifth Amendment to the United States Constitution provides that “No person shall . . . be deprived of life, liberty, or property, without due process of law.”

201. The Challenged Provision unconstitutionally delegates lawmaking authority to private actors—the website owners who draft terms of service. These private actors, and not any democratically accountable government entity, unilaterally determine which conduct is prohibited. The Challenged Provision does not place limits on what website owners may designate to be prohibited—and therefore, criminal—conduct.

202. The Challenged Provision violates the Due Process Clause of the Fifth Amendment.

Prayer for Relief

Plaintiffs respectfully request a judgment:

1. Declaring that the challenged provision, 18 U.S.C. § 1030(a)(2)(C), on its face and as applied to Plaintiffs, violates—
 - a. the Free Speech and Free Press Clauses of the First Amendment to the U.S. Constitution;
 - b. the Petition Clause of the First Amendment to the U.S. Constitution; and
 - c. the Due Process Clause of the Fifth Amendment to the U.S. Constitution;

2. Permanently enjoining the Defendant Attorney General, as well as her officers, agents, employees, attorneys, and all other persons in active concert or participation with her, from enforcing 18 U.S.C. § 1030(a)(2)(C);
3. Awarding Plaintiffs attorneys' fees and costs under the Equal Access to Justice Act, 28 U.S.C. § 2412; and
4. Awarding such other and further relief as this Court deems just and proper.

Dated: June 29, 2016

Respectfully submitted,

Esha Bhandari*
Rachel Goodman*
American Civil Liberties Union
Foundation
125 Broad St., 18th Floor
New York, NY 10004
Tel: 212-549-2500
Fax: 212-549-2654
ebhandari@aclu.org
rgoodman@aclu.org
**Application for admission pro hac vice
forthcoming*

/s/ Arthur B. Spitzer
Arthur B. Spitzer (D.C. Bar No. 235960)
artspitzer@aclu-nca.org
Scott Michelman** (D.C. Bar No. 1006945)
scott@aclu-nca.org
American Civil Liberties Union
of the Nation's Capital
4301 Connecticut Avenue, N.W., Suite 434
Washington, D.C. 20008
Tel: 202-457-0800
Fax: 202-457-0805
***Application for admission to this Court
pending*

Attorneys for Plaintiffs