Data Journalism and the Computer Fraud and Abuse Act: Tips for Moving Forward in an Uncertain Landscape

Esha Bhandari
Speech, Privacy & Technology Project,
American Civil Liberties Union
New York, NY
ebhandari@aclu.org

Rachel Goodman
Racial Justice Program,
American Civil Liberties Union
New York, NY
rgoodman@aclu.org

ABSTRACT

In this paper, we explore the implications of the Computer Fraud and Abuse Act ("CFAA") for journalists seeking to investigate websites and other publicly-available online resources. We describe current ACLU litigation seeking to protect these kinds of investigations and suggest best practices for avoiding CFAA liability.

1. INTRODUCTION

We need data journalists to inform public debate about online business practices, just as investigative journalists have long done in the offline world. We need robust journalism to help us make sense of the effects of these practices on civil rights, privacy, and consumer protections, among other things. It is especially crucial to investigate business practices as online experiences become ever more personalized and targeted. Users need to understand how online experiences are tailored specifically for them, possibly violating civil rights laws. For example, is a housing website suggesting the same homes to home-seekers of different races, or might it be illegally steering some users toward neighborhoods where, demographically speaking, they "belong"? With real-world transactions increasingly mediated through online platforms, the public should be able to learn how those systems determine outcomes for users, like where to send a limited number of taxis or who gets to see certain job ads. But because these decisions are often automated by code that is proprietary and hidden from public view, testing the outcomes produced by these algorithms is often the best available way to understand them.

Data journalists, along with academic researchers, are taking up this challenge. Unfortunately, they conduct these investigations in the shadow of a federal criminal statute called the Computer Fraud and Abuse Act ("CFAA") that perversely grants businesses that operate online the power to shut down any testing of their practices they don't like, regardless of the methods used. Intended to punish malicious hacking, the CFAA contains broad and vague language making it a crime to access a website in a manner that "exceeds authorized access." This provision has been interpreted to prohibit an individual from visiting a website in a manner that violates the website's terms of service.[1] However, common website terms of service prohibit activities like scraping publicly available information, creating multiple accounts, or providing false information, even though these activities are often necessary for robust testing, including the kind of testing that would uncover discrimination on the internet.

The government has never prosecuted a researcher or journalist for this kind of investigation—although it has prosecuted at least two people for violating terms of service in other contexts. The bad news is that the very existence of the law deters some journalists from conducting investigations that they would otherwise perform, and journalists who proceed may alter their methods in a way that makes the research less valuable. Additionally, the CFAA provides for civil liability in certain circumstances and, although journalists acting responsibly would have very strong arguments that they are not liable, the threat of being sued by the target of the investigation is, practically speaking, another deterrent they must deal with.

Last year, the ACLU filed a lawsuit on behalf of a group of academic researchers and a media organization seeking to remove the barrier posed by the CFAA's overbroad criminal prohibitions, and raising constitutional claims, including under the First Amendment. This paper will describe in detail that lawsuit and the problem that led us to file it. It will describe the ways in which this legal landscape may evolve. Finally, it will suggest methods for data journalists to protect themselves from CFAA liability when investigating online platforms.

A caveat: this paper focuses on data journalists' liability under the CFAA. It does not explore in any depth the various tort and contract law claims that the target of an investigation might seek to bring against a data journalist, although some information provided here will be relevant in that context, too. Such laws vary from state to state and will apply differently to various kinds of investigations. As a result, we encourage journalists to consult with legal counsel about their particular circumstances. We hope, however, that this paper helps journalists emerge from those consultations feeling confident that they can safely conduct these investigations, which are vitally important to protecting civil rights and to our public conversations about the use of technology more broadly.

2. THE PROBLEM

2.1 The Need for Testing

More and more of our economic, social, and political lives are mediated through algorithmic decisionmaking or machine learning. This includes transactions involving the core social goods covered by the civil rights laws, like housing, credit, and employment. Simultaneously, actions on the internet have lost the veil of anonymity as cookies and other tracking technologies allow websites to access all kinds of information about users. This information makes behavioral targeting possible, meaning that advertisers and websites can steer different individuals toward different products (or homes or credit offers or jobs), display different prices for the same products, or display different advertisements.[2]

The algorithms that determine who sees which product, or which price, or which advertisement, are proprietary and therefore exist in a kind of black box.[3] Moreover, these algorithms are often so complex that a sophisticated computer scientist examining the relevant code might not be able to fully explain which inputs lead to particular outputs. Some algorithms evolve constantly as they acquire new data. Accordingly, although many groups advocate for various forms of increased algorithmic transparency, a consensus exists that a crucial way to determine whether people are experiencing discrimination that would be prohibited by the civil rights laws is via outcomes-based audit testing.[4] Audit testing allows researchers to uncover disparate impact—namely, whether people are being treated differently on the basis of their protected class status (such as race or gender).

For more than three decades, the offline equivalent of audit testing has been a core part of enforcing the Fair Housing Act ("FHA"). Paired testing has become the standard procedure for determining whether a housing provider is discriminating because it is nearly impossible for an individual to know whether she has been discriminated against without knowing about the experiences of other prospective renters or buyers. In a paired test, two people who differ only in their race or gender or disability status (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 landlord or real estate agency. Court decisions, a federal statute, and various regulations provide ample support for such testing, and make clear that testing is central to achieving the goals of the FHA. Given "the enormity of the task of assuring fair housing" and the importance of "'private attorneys general'" in that task,[5] the Supreme Court has held that individual "testers" may feign interest in purchasing or renting a property in order to investigate potential violations of the FHA and challenge these violations in court.[6] Testing has also played a role in the enforcement of other anti-discrimination laws. 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, for example. In an in-person audit study, pairs of real testers apply for jobs, presenting equal credentials and comparing outcomes.

If anti-discrimination laws are to remain effective in the increasingly digital marketplace, online versions of this testing must be allowed and, in fact, encouraged. Yet online research designed to uncover discrimination—and many other forms of online investigation—will very often, if not always, violate terms of service ("ToS"). For example, research designed to reveal how particular algorithms operate often involves using automated technology to record the information that is publicly presented, or to access a website repeatedly—generally by creating false or artificial user profiles—in order to examine whether and how platforms respond differently to different users. Many ToS prohibit the use of automated technology and the creation of fictitious user profiles, and some even aim to prohibit "disparaging" comments about the website or business. And, as detailed below, such terms of service violations may form the basis for liability under the CFAA.

2.2 CFAA Liability for Violating Terms of Service

When data journalists access a website in a manner that violates its ToS, their potential liability resides primarily in the provision of the CFAA that makes it a misdemeanor to obtain information from a protected computer where a user "intentionally accesses a computer without authorization or exceeds authorized access."[7] Courts interpret the broad statutory definition of "protected computer" to mean any computer or other device connected to the internet,[8] and the law does not require that the accessor intended to cause harm or actually inflicted harm before being held liable. The basic penalty for a first violation of this provision is a one-year maximum prison sentence and a fine, which can rise to ten years for a repeat offense.[9]

"Exceeds authorized access" is defined in the CFAA to mean "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."[10] Some courts, and some federal prosecutors, have interpreted this vague language to encompass access in violation of written conditions—that is, terms of service. In one case, a woman who had impersonated a young man on MySpace and, using that profile, tormented a teenager until she committed suicide, was prosecuted not only for cyberbullying but for creating the false profile in violation of MySpace ToS.[11] A judge ultimately threw out the conviction, but the Department of Justice has not renounced prosecution based on ToS violations.[12]

Although numerous courts have rejected the notion that ToS violations constitute CFAA crimes,[13] so long as the statute remains on the books and the broader and more aggressive interpretation has not been definitively rejected by the Supreme Court, the worrisome possibility of prosecution for terms of service liability remains.

2.2.1 Civil Liability

Private companies can also sue for damages for violations of the CFAA,[14] meaning that there is also a theoretical risk of civil liability for data journalists who perform tests in violation of ToS. However, the statute enumerates several narrow circumstances in which civil liability can exist.[15] For data journalists who do not test federal government computers or cause actual damage to computers, the only plausible argument for civil liability would be that the investigation had caused $5,000 or more in loss during a 1-year period.[16] While it is possible that an investigation could cause this kind of damage by, for example, creating a load on targeted servers of a magnitude that would slow or disable them, thus interfering with business, investigations can usually be designed to avoid this undesirable consequence. Moreover, to the extent that the target of investigation might want to argue that negative publicity stemming from publication of truthful findings is the type of "harm" covered by the law, there are strong arguments to rebut this claim. A reading of the statute defining "harm" to include reputational effects of publication would raise serious First Amendment concerns, and similar arguments have been rejected by courts considering harm or damage in other contexts.

As a result, it is very unlikely that damages could ultimately be assessed against a data journalist or her employer for conducting the most common types of online audit tests.

2.3 *Sandvig v. Lynch* – The Litigation

Last year, the ACLU filed a lawsuit asking a federal court to declare the "exceeds authorized access" provision of the CFAA to be unconstitutional.[16] We filed the lawsuit on behalf of academic researchers and a media organization so that they can conduct tests to investigate online discrimination without fear of liability under the CFAA. We argue that the CFAA violates the First and Fifth Amendments of the U.S. Constitution because it is overbroad and vague and because it delegates too much authority

to private parties to determine what conduct is criminal. We also argue that the CFAA limits everyone, including academics and journalists, from engaging in the activity necessary to understand and speak about online discrimination—such as recording publicly available information and publishing findings. These constitutionally protected activities become potential criminal violations of the CFAA whenever website owners decide to prohibit them through ToS.

The federal government moved to dismiss the lawsuit, claiming that these investigative activities are not covered by the First Amendment and that our plaintiffs are not at risk of being prosecuted under the CFAA and therefore do not have standing to sue. In support of its argument, the government publicly released a memorandum detailing the factors that the Department of Justice considers when deciding whether to prosecute a CFAA violation.[17] The memorandum suggests that the government does not consider prosecution of terms of service violations to be a high priority, noting that "if the defendant exceeded authorized access solely by violating an access restriction contained in a contractual agreement or term of service with an Internet service provider or website, federal prosecution may not be warranted." However, because it does not disavow terms of service prosecutions, it does not remove the fear of prosecution that gives the plaintiffs standing to sue. The court has not yet ruled on the motion to dismiss.

While the *Sandvig v. Lynch* case is pending, other developments in the law may affect journalists' liability under the CFAA, as described in the section below.

3. EVOLVING LANDSCAPE – ISSUES TO WATCH

3.1 Supreme Court Cert Petitions Pending

The Supreme Court may soon weigh in on aspects of the CFAA that affect liability for testing online. It is currently considering whether to take up two cases arising out of the Ninth Circuit Court of Appeals, which address the provision of the CFAA prohibiting access to a protected computer "without authorization."[18] Although this provision has not been interpreted to prohibit violations of ToS alone, the Supreme Court's view of the types of behavior that constitute access "without authorization" could have implications for online journalism and testing. Neither case involves research or journalism. One is about whether a former employee can be held liable for accessing the employer's computers using login credentials voluntarily given by another employee, and the other is about whether a competitor to Facebook could be liable for accessing Facebook accounts with the permission of the individual account users. Nonetheless, resolving these cases would likely involve addressing the extent to which CFAA liability depends on whether you know that the owner of a website or computer wants to prohibit you from accessing that website or computer. This issue is especially relevant to journalists when platform owners have directly told them not to conduct particular research.

3.2 Reform Efforts in Congress

While courts continue to grapple with the interpretation of the CFAA and how it should be applied in particular cases, there is a growing recognition in Congress that the CFAA must be reformed, particularly given vast changes in the internet landscape since it was enacted in the 1980s. Various amendments have been proposed in the past few years, most notably "Aaron's Law," which aims to address some of the overbreadth and vagueness problems in the CFAA.[19] Legislative reform of the CFAA is critical, given that court decisions may not resolve every problematic aspect of the CFAA. Reform efforts must consider the effect the CFAA has on researchers and journalists and the risk that civil rights laws will be under-enforced online because businesses try to block the testing necessary for that enforcement.

4. TIPS FOR JOURNALISTS

This section suggests best practices for data journalists who wish to engage in audit testing online while insulating themselves against both criminal and civil liability under the CFAA. Of course, incorporating these considerations cannot guarantee that you will never face litigation or the threat of it. It is up to individual journalists and media organizations to assess their tolerance for risk. It is our hope, however, that this vitally important work will continue given that it is a strong matter of public concern.

4.1 First Do No Harm

Perhaps most obviously, data journalists looking to avoid CFAA liability and other forms of civil liability should carefully tailor their investigations to avoid placing too much stress on the target's infrastructure. The goal is to ensure, to the greatest extent possible, that the targeted computers or servers continue to function as they would have in the absence of the investigation. Designing an investigation in this way will help guard against allegations that the investigation caused damage to the machines and to the regular business operations of the targeted entity.

In practice, this means considering the overall capacity of a target machine or service and ensuring that any requests generated by bots or scrapers involved in the investigation make up a minimal share of that capacity at any given time. It may mean, for example, designing software to make a small number of requests repeatedly over a long period of time, rather than overwhelming a server by running all of the requests at once. Journalists should also consider running bots and scrapers at off-hours, when servers are not likely to be experiencing much traffic, though this may not be possible for robust investigation into some services (e.g., trip or route planners) which are highly sensitive to time-of-use. Finally, investigations that trigger external events (such as hailing a car service, reserving lodging, or ordering delivery of groceries) to see how the system behaves should be limited in scope, particularly if the actions are taken with the intent of prompt cancellation of the service requested.

4.2 Does Fear of Negative Publicity Protect You?

Imagine: a data journalist working for a major publication conducts an investigation that reveals that a platform operated by a large and publicly-traded company systematically disadvantages women or people of color in some way. When the platform gets wind of the investigation, it sues the journalist and the publication claiming damages from the test. How does the company look to the public when news of this retaliatory suit gets out?

Few major companies would relish the prospect of finding out. In recent years, many technology companies have shown themselves to be sensitive to allegations of discrimination (or even inaction in the face of discrimination), and, more broadly, to publicity that makes them look like bad actors. This is especially true when their own employees are angered by corporate action. This sensitivity offers data journalists some protection. How much will depend on the footprint of the journalist and publication involved; the size and corporate culture of the target; and the extent to which its business is public-facing and relies on the decisions of individual

consumers. It will likely also depend on the particulars of the investigation—the more newsworthy the topic, the more protection a journalist may derive. For example, an investigation into gender discrimination in job recruiting may generate widespread interest (and thus protection via public attention). An investigation that can be tied to efforts to enforce civil rights laws (or other laws) may also enjoy greater protection from retaliatory civil suits.

4.3 Consider Informing the Investigated Entity

One possibility researchers should consider is informing the investigated entity of their plans and receiving permission. In a best-case scenario, the targeted entity will grant permission, precluding any argument that testing activities violated the CFAA's authorization provisions. However, if the targeted entity refuses permission, a researcher may find herself in a worse legal position than before. The Ninth Circuit Court of Appeals, in the Facebook case that may go to the Supreme Court, considered it a significant factor in imposing CFAA liability that Facebook had sent the competitor a cease-and-desist letter, putting it on notice that its access to Facebook accounts was unauthorized by Facebook.[20] While that case arose under a different provision of the CFAA than the provision that has been interpreted to cover violations of ToS, it is nonetheless important for researchers to consider when evaluating whether to seek explicit permission and risk being denied. (There may be, of course, other downsides to seeking permission, including the possibility that the targeted entity changes its behavior in advance of the proposed research.)

4.4 Defense Based on Civil Rights Enforcement

Should a data journalist conducting research into algorithmic discrimination find herself in court, defending against the allegation that she should be held liable for accessing or copying or publishing information that she obtained through some form of falsity or deception, she can defend herself by arguing that the testing she engaged in was merely the online equivalent of offline testing long approved by the courts.

Courts recognize, in the context of fair housing, that testers are necessary for enforcement, even though they are not genuinely interested in the housing they claim to seek during the test.[21] Courts even acknowledge explicitly that there is deception involved in testing, and nonetheless permit it. As one appellate court put it, "It is surely regrettable that testers must mislead commercial landlords and home owners as to their real intentions. . . . Nonetheless, we have long recognized that this requirement of deception was a relatively small price to pay to defeat racial discrimination. The evidence produced by testers . . . is a major resource in society's continuing struggle to eliminate the subtle but deadly poison of racial discrimination."[22] Congress passed a statute ensuring that the federal government funds this kind of testing directly.[23] Testing has similarly been recognized by some courts as a vital part of the enforcement of anti-discrimination laws in employment.[24]

The more closely an online audit test resembles these offline tests, the more persuasive this argument will likely be to a court. The argument is strongest with respect to housing-related investigations, slightly less strong where it concerns employment-related issues, and still worth raising any time an investigation concerns discrimination prohibited under anti-discrimination laws.

5. CONCLUSION

Data journalists have a vital role to play in helping the public understand the ways in which algorithms sort and treat people differently as they browse the internet, through processes that have huge implications for civil rights. Audit testing of online platforms is crucial to discovering, and hence fixing, discrimination online, and the CFAA ought not to present a barrier to journalists interested in conducting that testing. The authors of this paper encourage journalists wrestling with these issues both to consult with legal counsel and to reach out to us—we would be pleased to continue this conversation with respect to particular circumstances.

6. REFERENCES

[1] 18 U.S.C. § 1030(a)(2)(C).

[2] *See, e.g.,* Anniko Hannak et al., *Measuring Price Discrimination and Steering on E-Commerce Websites*, *available at* http://www.ccs.neu.edu/home/cbw/pdf/imc151-hannak.pdf (2014); Jennifer Valentino-Devries et al., *Websites Vary Prices, Deals Based on Users' Information*, Wall Street Journal, Dec. 24, 2012, http://www.wsj.com/articles/SB10001424127887323777204578189391813881534.

[3] *See* Frank Pasquale, *The Black Box Society*: *The Secret Algorithms that Control Money and Information*, Harvard Univ. Press (2015).

[4] Christian Sandvig et al., *Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms* (2014), *available at* http://www-personal.umich.edu/~csandvig/research/Auditing%20Algorithms%20--%20Sandvig%20--%20ICA%202014%20Data%20and%20Discrimination%20Preconference.pdf.

[5] *Trafficante v. Metro. Life Ins. Co.*, 409 U.S. 205, 211 (1972).

[6] *Havens Realty Corp v. Coleman*, 455 U.S. 363, 373 (1982).

[7] 18 U.S.C. § 1030(a)(2).

[8] 18 U.S.C. § 1030(e)(2)(B); *United States v. Trotter*, 478 F.3d 918, 921 (8th Cir. 2007).`

[9] 18 U.S.C. § 1030(c)(2)(A), (C).

[10] 18 U.S.C. § 1030(e)(6).

[11] *United States v. Drew*, 259 F.R.D. 449 (C.D. Cal. 2009).

[12] Dep't of Justice, Office of the Atty. Gen., Intake and Charging Policy for Computer Crime Matters (Sept. 11, 2014), *available at* https://www.aclu.org/sites/default/files/field_document/15-1_ex_to_mtd_reply_-_charging_memo.pdf; 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.

[13] *See, e.g.*, *United States v. Valle*, 807 F.3d 508, 527 (2d Cir. 2015); *WEC Carolina Energy Sols. LLC v. Miller*, 687 F.3d 199, 206 (4th Cir. 2012); *United States v. Nosal*, 676 F.3d 854, 857 (9th Cir. 2012).

[14] 18 U.S.C. § 1030(g).

[15] *Id.* (cross-referencing 18 U.S.C. § 1030(c)(4)(A)(i)).

[16] Complaint, *Sandvig v. Lynch*, No. 16 Civ. 01368 (D.D.C., filed June 29, 2016), *available at*

https://www.aclu.org/sites/default/files/field_document/cfaa_complaint_0.pdf.

[17] Dep't of Justice, Office of the Atty. Gen., Intake and Charging Policy for Computer Crime Matters (Sept. 11, 2014), *available at* https://www.aclu.org/legal-document/sandvig-v-lynch-office-us-attorney-general-intake-and-charging-policy-computer-crime.

[18] *See Nosal v. United States*, No. 16-1344 (9th Cir.); *Power Ventures, Inc.*, *v. Facebook, Inc*., No. 16-1105 (9th Cir).

[19] Zoe Lofgren & Ron Wyden, *Introducing Aaron's Law, a Desperately Needed Reform of the Computer Fraud and Abuse Act*, Wired, June 20, 2013, https://www.wired.com/2013/06/aarons-law-is-finally-here/.

[20] *See Power Ventures, Inc.*, *v. Facebook, Inc*., 844 F.3d 1058, 1067 (9th Cir. 2016).

[21] *Havens Realty Corp v. Coleman*, 455 U.S. 363, 373 (1982).

[22] *Richardson v. Howard*, 712 F.2d 319, 321 (7th Cir. 1983).

[23] 42 U.S.C. § 3616a (b)(2)(A), (C).

[24] *Kyles v. J.K. Guardian Sec. Servs., Inc.*, 222 F.3d 289 (7th Cir. 2000).