The Color of Surveillance: Examining the Complicity of Social Media Companies In Facilitating Law Enforcement Surveillance of Civil Rights Activism

Meghan Koushik

INTRODUCTION

The #BlackLivesMatter movement has capitalized on social media in a way few other popular movements before it were able to do. In previous decades, civil rights activists seeking to transmit the realities of being black in America—from lynching, to hate crimes, to burning crosses—relied on an extensive—and slow-- infrastructure of line-to-line communication to get their stories from the Jim Crow South to a national audience.[1] In the Black Lives Matter age, the proliferation of social media platforms has caused a seismic shift in the infrastructure required to give racism and white supremacy a nationwide spotlight. When Alton Sterling was shot and killed after an encounter with police on the streets of Baton Rouge, Louisiana, an anti-violence group nearby captured the full encounter on video, showing Sterling being wrestled down, then shot and killed while pinned on the ground by two white police officers. The group eventually posted the video to Facebook and Twitter, disputing the police's account that Sterling was a threat to their safety.[2] Less than a day later, Philando Castile, a second African-American man was shot and killed by police during a traffic stop. Fearing for her own life, his girlfriend in the seat next to him streamed the aftermath of the shooting through Facebook Live, allowing millions of Americans to

[1] See Bijan Stephen, *Social Media Helps Black Lives Matter Fight The Power,* WIRED (Nov. 2015), https://www.wired.com/2015/10/how-black-lives-matter-uses-social-media-to-fight-the-power/ (noting the prolonged amount of time it took to communicate information through use of WATS lines and reports as mechanisms of information transmission in the Civil Rights Era)

[2] *See, e.g.,* Travis M. Andrews, *The Story Behind The Filming of the Fatal Baton Rouge Police Shooting,* WASH. POST (Jul. 7, 2016) https://www.washingtonpost.com/news/morning-mix/wp/2016/07/07/the-story-behind-the-filming-of-the-fatal-baton-rouge-police-shooting/?noredirect=on&utm_term=.f5e0bc6213d4; Wesley Lowery et al., *Video Captures White Baton Rouge Police Officer Fatally Shooting Black Man, Sparking Outrage,* WASH. POST (Jul. 6, 2016), https://www.washingtonpost.com/news/morning-mix/wp/2016/07/06/video-captures-white-baton-rouge-police-officer-fatally-shooting-black-man-sparking-outrage/?utm_term=.d10d3f027511.

share her anguish in real-time as Castile bled to death.[3] Both killings sparked mass outrage nationwide, causing thousands to march in protest nationwide and drawing comments from then-President Obama as well as a host of national leaders.[4] These are just two examples of how social media platforms are changing the landscape of social justice movements. Increasingly, people of color across America are using Facebook and Instagram to live-stream incidents of police brutality; relying on GroupMe, WhatsApp and Signal to coordinate national-level campaigns overnight; and finding themselves an audience of millions to communicate with in real-time on platforms like Twitter.

At the same time, even as social movements increasingly rely on social media for organizing and activism, law enforcement agencies are relying on them to constrain these movements. People put massive quantities of personal information on social media—everything from postings about their political views, to their precise locations and movement. Similarly, the networking aspect of social media means that outside observers can glean significant amounts of information not just on individuals, but on broader groups and even movements, by tracking patterns of user engagement, online "groups" and event pages, and even hashtags. Moreover, it seems like almost everyone is online. Nearly a quarter of all U.S. citizens are monthly active Twitter users,[5] while Facebook currently boasts 2.27 billion monthly active users.[6]

---

[3] Chelsea Bailey, *Philando Castile Shooting: Girlfriend Testified She Began FB Live Because She Feared For Her Own Life,* NBC NEWS (June 6, 2017), https://www.nbcnews.com/news/us-news/philando-castile-killing-girlfriend-testifies-she-began-fb-live-because-n768916.

[4] Mitch Smith, *Minnesota Officer Acquitted in Shooting of Philando Castile,* N.Y. TIMES (June 16, 2017), https://www.nytimes.com/2017/06/16/us/police-shooting-trial-philando-castile.html (quoting the governor of Minnesota, Mark Dayton, who asked aloud: "Would this have happened if the driver were white, if the passengers were white?")

[5] Kit Smith, *58 Incredible and Interesting Twitter Stats and Statistics,* BRANDWATCH (Jan. 3, 2019), https://www.brandwatch.com/blog/twitter-stats-and-statistics/

[6] Jason Abbruzzese, *Facebook Hits 2.27 Billion Monthly Active Users as Earnings Stabilize,* NBC NEWS (Oct. 30, 2018), https://www.nbcnews.com/tech/tech-news/facebook-hits-2-27-billion-monthly-active-users-earnings-stabilize-n926391

Law enforcement has been quick to capitalize on this trove of information. A 2016 study found that at least 70% of police agencies reported using social media to conduct intelligence gathering for investigations and 72% used it to "monitor public sentiments."[7] At least 60% have contacted a social media company like Facebook or Twitter to request online information to use as evidence in a legal proceeding.[8] In the first half of 2017, the top six internet companies—Google, Facebook, Microsoft, Apple, Oath (formerly Yahoo!) and Twitter-- received over 130,000 requests for user information from law enforcement, marking almost a 50% increase since 2014.[9]

The rise of social media surveillance raises significant questions about its implications for constitutional rights. Surveillance of social and political movements is not new—even as Martin Luther King Jr. is celebrated as a civil rights leader today, in his lifetime, he was the subject of relentless surveillance and scrutiny by American law enforcement and intelligence, as were many others involved in the civil rights movement.[10] However, in previous decades, where law enforcement relied on tools like wiretapping and physical surveillance of activists and organizers, they faced both resource and legal constraints that limited the scope and duration of such surveillance. As technology evolves, it becomes increasingly cheaper and less labor-intensive to expand the surveillance net. The majority of police departments are conducting social media surveillance without clear internal guidelines, restrictions, or oversight.[11] As with previous eras, such

---

[7] KIDEUK KIM ET AL., URBAN INSTITUTE, 2016 LAW ENFORCEMENT USE OF SOCIAL MEDIA 3 (2017), https://www.urban.org/sites/default/files/publication/88661/2016-law-enforcement-use-of-social-media-survey_5.pdf.

[8] *Id.* at 4.

[9] WILLIAM A. CARTER & JENNIFER DASKAL, CTR. FOR STRATEGIC & INT'L STUDIES, LOW HANGING FRUIT: EVIDENCE-BASED SOLUTIONS TO THE DIGITAL EVIDENCE CHALLENGE 17 (2018), https://csis-prod.s3.amazonaws.com/s3fs-public/publication/180725_Carter_DigitalEvidence.pdf?tAGR_DvxRdp0RspiGYNGcGKTUjrGY3rN (showing approximately 87,000 requests to the same companies in 2014; compared to 130,000 in 2017).

[10] Alvaro M. Bedoya, *The Color of Surveillance,* SLATE (Jan. 18, 2016), https://slate.com/technology/2016/01/what-the-fbis-surveillance-of-martin-luther-king-says-about-modern-spying.html; Jeanne Theoharis, *Comey Says FBI's Surveillance of MLK Was "Shameful,"—But Comey's FBI Targeted Black Activists and Muslim Communities Anyway,* INTERCEPT (Apr. 24, 2018), https://theintercept.com/2018/04/24/james-comey-mlk-martin-luther-king-surveillance-muslims/

[11] LEXISNEXIS, SOCIAL MEDIA USE IN LAW ENFORCEMENT 2 (2014), https://risk.lexisnexis.com/insights-resources/infographic/law-enforcement-usage-of-social-media-for-investigations-infographic (finding 52% of polled police departments had no formal processes in place for using social media).

surveillance disproportionately impacts communities of color, particularly African-Americans and Muslims.[12] And particularly where social media surveillance is concerned, few, if any, courts have interpreted the Fourth Amendment expansively enough to decisively curb (or even constrain) such surveillance. [13]

This paper will therefore address the other key player in this equation with the capacity to place limits on how law enforcement utilizes social media for surveillance—the social media companies themselves. Fundamentally, this paper will demonstrate that social media companies themselves can play critical roles in regulating the use of their data and platform for surveillance, both by making affirmative changes to their own guidelines and policies, and aggressively enforcing violations of these terms. This paper will proceed in four parts. First, it will provide an overview of four key ways in which law enforcement have utilized social media to conduct surveillance. Second, the paper will provide a brief overview of the constitutional rights implicated by such surveillance and discuss why the legal system may not be the most effective mechanism to challenge such surveillance in the short-term. Third, the paper will turn to the platforms themselves, and perform a comparative analysis of their privacy and data retention policies, as well as their law enforcement guidelines. This analysis will inform the final section, which will provide recommendations for how platforms aiming to curb the use of their product and data for surveillance purposes can strengthen their own policies and guidelines in order to prevent such uses.

PART I: HOW DO POLICE USE SOCIAL MEDIA FOR SURVEILLANCE?

This section will provide an overview of four key social media surveillance techniques employed by law enforcement: first, searches and indexes of publicly available information; second,

[12] *See infra* Part I.
[13] *See infra* Part II.

the creation and use of fake or fraudulent accounts by law enforcement to technically or socially circumvent site privacy measures; third, formal legal requests for user account information and activity; and finally, reliance on data-mining software like Geofeedia that can aggregate and analyze social media feeds to provide real-time monitoring of movements. It will further contextualize these mechanisms by discussing real-life instances where law enforcement agencies have relied on such tools to engage in surveillance and censorship of civil rights activists and movements. This analysis ultimately aims to demonstrate the chilling effect such social media imposes on communities of color and the movements for change they seek to lead.

A. Searches and Tracking of Publicly Available Information

The simplest—and likely most common means of social media surveillance is extrapolating information from publicly posted content. Despite a range of privacy controls offered by major social media companies,[14] a large number of users on platforms like Twitter choose to keep their accounts public, providing a rich trove of publicly-available information for law enforcement to access, from pictures, to public posts, to even regularly-updated location data.[15] Law enforcement will often push back with the assertion that in some instances, such monitoring has led to immediate payoffs. There have been a number of cases—both in the US and internationally—where criminals have posted direct evidence of their crimes on social media platforms, or even used real-time services like Facebook Live to live-stream their criminal activities.[16]

---

[14] Social media platforms offer a range of options for user privacy. For instance, Instagram and Twitter require user accounts to be either fully viewable to the public, or "private," wherein only approved followers can view posted content, whereas Facebook takes a somewhat more nuanced approach by allowing users to change approved audiences for each piece of content posted.

[15] *Twitter Data Analysis: An Investor's Perspective,* TECHCRUNCH (Oct. 5, 2009), https://techcrunch.com/2009/10/05/twitter-data-analysis-an-investors-perspective-2/ (finding only around 10% of Twitter accounts are set to private).

[16] Olivia Solon, *Why a Rising Number of Criminals Are Using Facebook Live to Film Their Acts,* GUARDIAN (Jan. 27, 2017), https://www.theguardian.com/technology/2017/jan/27/rising-numbers-of-criminals-are-using-facebook-to-document-their-crimes.

Perhaps more perniciously, however, such social media surveillance has become a key tool both for police departments' anti-gang efforts; as well as to law enforcement monitoring of large-scale events and movements, such as protests or rallies. Critics have raised alarms over such surveillance and its disproportionate impact on minority communities, with some even calling social media surveillance "stop-and-frisk online."[17] Such critics claim that law enforcement uses evidence from social media surveillance not to prove criminal wrongdoing, but rather to implicate black and brown teenagers on conspiracy charges based on their social media interactions with other individuals deemed to be affiliated with gangs—even when they may not have engaged in actual criminal conduct themselves. Faced with the prospect of extended jail time on conspiracy charges, many will plead guilty to lesser sentences, ultimately furthering the cycle of incarceration that already disproportionately impacts minority communities.

The New York City Police Department ("NYPD")'s use of social media monitoring has been one prominent example of this use. The NYPD has been extremely aggressive in recent years about using social media monitoring as a tool in investigations and evidence-gathering, particularly in the context of gang-related criminal activity.[18] The NYPD's Juvenile Justice Division "focuses on analyzing social networking by local youth gangs and neighborhood crews," while a specialized group within the Intelligence Division watches social media for information on "large-scale events and criminal activity."[19] In the context of gang activity, the NYPD frequently makes determinations about an individual's gang membership based on their social media activity—being photographed with other gang members, "liking" their posts, or wearing colors or other paraphernalia associated

---

[17] *See, e.g.,* Desmond Upton Patton et al., *Stop and Frisk Online: Theorizing Everyday Racism in Digital Policing,* SOCIAL MEDIA + SOCIETY (2017).

[18] Sara Robinson, *When a Facebook Like Lands You in Jail,* BRENNAN CTR. FOR JUSTICE (Jul. 6, 2018), https://www.brennancenter.org/blog/when-facebook-lands-you-jail.

[19] POLICE EXECUTIVE. RESEARCH FORUM, SOCIAL MEDIA AND TACTICAL CONSIDERATIONS FOR LAW ENFORCEMENT 13, https://www.policeforum.org/assets/docs/Free_Online_Documents/Technology/social%20media%20and%20tactical%20considerations%20for%20law%20enforcement%202013.pdf.

with a gang, for instance, are treated by the NYPD as evidence of gang membership. Such individuals are then marked for inclusion in the NYPD's internal gang database.[20] If arrested, individuals deemed "gang members" in this manner generally face much higher barriers to obtaining bail and will often be subjected to harsher outcomes during sentencing.[21]

The NYPD has also used social media evidence directly as a basis for indicting and prosecuting suspected gang members. Under its "Operation Crew Cut," which "combin[es] a focus on crews and social media in an effort to curb gun violence,"[22] the NYPD has conducted a series of "gang raids" to sweep up and indict suspected gang members, many of whom later find that their past social media postings appear in their indictments. For instance, in 2014, over a hundred Harlem residents—many under the age of 20—were arrested and indicted in the largest gang raid in New York City history.[23] The majority were ultimately indicted on gang-related conspiracy charges.[24] Reportedly, "to build the case for the raid, police had begun social media surveillance of children well before they had built up a serious criminal record."[25] Media outlets reported that "for over four years, prosecutors and police investigators combed more than a million social media pages"[26] to derive evidence, and "[t]he word 'Facebook' appear[ed] more than three hundred times in [their] indictments."[27] At least ninety-three of the defendants ultimately took plea deals to avoid the lengthy sentences they would have received if found guilty on gang conspiracy charges.[28] In another well-

---

[20] Robinson, *When a Facebook Like Lands You in Jail, supra* note 18.
[21] There can also be immigration related consequences: the NYPD has worked with ICE on "gang takedowns," essentially mass raids intended to sweep up hundreds of purported gang members in a single swoop. *See* George Joseph, *Has Gang Policing Replaced Stop-and-Frisk?* CITYLAB (Feb. 28, 2017), https://www.citylab.com/equity/2017/02/has-gang-policing-replaced-stop-and-frisk/517572/.
[22] Ben Popper, *How the NYPD is Using Social Media to Put Harlem Teenagers Behind Bars,* VERGE (Dec. 10, 2014), https://www.theverge.com/2014/12/10/7341077/nypd-harlem-crews-social-media-rikers-prison.
[23] *Id.*
[24] Josmar Trujillo, *Gangbusters: Law Enforcement's Dubious Means to Police Public Housing Residents,* TRUTHOUT (May 6, 2016), https://truthout.org/articles/gangbusters-law-enforcement-s-dubious-means-to-police-public-housing-residents/
[25] Popper, *How the NYPD is Using Social Media, supra* note 22.
[26] *Id.*
[27] *Id.*
[28] Trujillo, *Gangbusters: Law Enforcement's Dubious Means, supra* note 24.

documented instance, Asheem Henry, a former gang member, was arrested on conspiracy charges largely based on social media posts he had made as a teenager. Shortly afterwards, his younger brother Jelani was arrested after being wrongly accused of attempted murder. At his arraignment, the DA's office painted him as a "violent gang member" by introducing Jelani's social media activity as evidence, including pictures with members of his brother's "crew." As a result, Jelani was denied bail and incarcerated on Rikers Island for over two years while awaiting trial, including nine months in solitary confinement.[29] He was subsequently exonerated.[30] Similar cases have come to light around the country.[31]

The Philadelphia Police Department's Focused Deterrence program operates in a similar manner. In one case, a teenager was deemed a "gang member" based on photos and tweets, as well as an appearance in a friend's music video. Police later arrested him on suspicion that he had been involved in a shooting, and used his "gang affiliation" to argue against bail. He was later cleared of all charges, but only after being jailed for several months.[32] In another instance, after a second teenager tweeted lyrics from a rap song referencing a firearm, police raided his home, found a firearm in a separate bedroom, and slapped him with illegal gun possession charges. Police attempted to argue that the rap lyrics indicated that the gun belonged to the teenager, though the gun was not found in his room and there was no evidence that he had so much as touched it. The charges were later thrown out.[33]

---

[29] Popper, *How the NYPD is Using Social Media, supra* note 22.
[30] *Id.*
[31] For instance in California, two African-Americans filed a §1983 lawsuit against San Diego city and its police after an allegedly unlawful arrest. Police affidavits supporting their arrest warrant cited the pair's social media activity, including appearing in pictures with known gang members and posting slang associated with a local gang, as justification for their arrest. Complaint, Duncan v. City of San Diego, CV-0052-BTM-MMD (S.D. Cal 2017), *available at* http://www.abajournal.com/files/WrongfulRapperArrest.pdf.
[32] At a hearing, the indicting officer reportedly dismissed a number of positive character references on the teenager, stating "I don't gather information from the community, and quite frankly, I don't care what other people's opinions are. We look at what they put on social media." Max Rivlin-Nadler, *How Philadelphia's Social-Media Driven Gang Policing Is Stealing Years from Young People,* The Appeal (Jan. 19, 2018), https://theappeal.org/how-philadelphias-social-media-driven-gang-policing-is-stealing-years-from-young-people-fa6a8dacead9/ .
[33] *Id.*

As such, what may seem like relatively benign social media surveillance has damning consequences for individuals. Evidence derived off social media can support broader conspiracy charges, elevating relatively low-level offenses into serious crimes with lengthy jail sentences. As in Jelani's case, individuals deemed "gang members" are treated far more harshly in the criminal system—unable to obtain bail, often incarcerated in harsher conditions, and if sentenced, usually subjected to more severe prison terms. Moreover, the sort of evidence deemed probative from such surveillance is prone to misinterpretation. The NYPD has used both indications of a user's diminished social media activity, as well as increased frequency in online posting, as "proof of criminality."[34] In such instances, individuals are painted as guilty whether they post too much or too little, raising doubts that any conclusive inferences can be drawn from either extreme. Similarly, as with Jelani's case, prosecutors have pointed to "liking" or otherwise engaging with alleged gang members' social media activity as evidence of gang participation. However, for many teenagers from these neighborhoods, how they are "perceived on the digital street" can be a survival mechanism rather than a signal of gang membership. Jelani explained in an interview that even for someone like him who was not involved in a gang or crew, actual gang members "are looking to see how you respond…[i]f you don't 'like' [their] post[s]…people are gonna ask you why."[35]

Critics deeming such surveillance "the new stop-and-frisk," point out that such activities tend to disproportionately impact low-income minority communities already bearing the brunt of overpolicing.[36] The majority of teenagers swept up in the NYPD's gang raids, for instance, are generally African-American and Latinx youths. Similarly, the NYPD's gang database is comprised of at least 95% people of color, even though such communities represent a far smaller proportional

[34] Trujillo, *Ganbusters: Law Enforcement's Dubious Means, supra* note 24.
[35] Popper, *How the NYPD is Using Social Media, supra* note 22.
[36] Rose Hackman, *Is The Online Surveillance of Black Teenagers The New Stop-and-Frisk?* Guardian (Apr. 23, 2015), https://www.theguardian.com/us-news/2015/apr/23/online-surveillance-black-teenagers-new-stop-and-frisk; *see also* Robinson, *When a Facebook Like Lands You in Jail, supra* note 18.

share of the city's overall population.[37] Moreover, the growing awareness that police are monitoring social media networks within these communities likely propagates a chilling effect, even for individuals who have no connection to criminal activity. Such individuals may feel additionally constrained from expressing criticism of the police, for instance, or responding to instances of police brutality, for fear that such posts might make them a target for surveillance, and ultimately come back to haunt them in future proceedings. The chilling effect of such surveillance will be explored in more detail at the end of this section.

B. Fake Accounts

The tactics above have generally relied on publicly-posted information. However, in recent years, a number of police departments have acknowledged using a version of "catfishing," or creating fake accounts on social media platforms, in order to access individuals' non-public postings. Not only does this provide law enforcement with far greater access to information, particularly on Facebook, this provides a means of circumventing the site's own privacy protections for users. According to the *New York Times,* under the NYPD's Juvenile Robbery Intervention Program, or J-RIP, police officers create "dummy" Facebook profiles, often featuring an "attractive teenage girl" and "send out 'friend requests' as bait to get beyond the social network's privacy settings."[38] Becoming Facebook "friends" naturally allowed NYPD detectives to see information that would otherwise be unavailable on a public profile. Detectives reported spending hours every day monitoring teenagers' online activity in this manner—tracking largely Black and Hispanic

[37] Robinson, *When a Facebook Like Lands You in Jail, supra* note 18.
[38] Wendy Ruderman, *To Stem Juvenile Robberies, Police Trail Youths Before The Crime,* N.Y. TIMES (Mar. 3, 2013), https://www.nytimes.com/2013/03/04/nyregion/to-stem-juvenile-robberies-police-trail-youths-before-the-crime.html.

teenagers—though department rules prevent them from directly engaging or communicating with the teenagers they tracked. [39]

Police have also used this tactic to surveil and infiltrate broader civil rights movements. In a lawsuit filed by the ACLU of Tennessee, for instance, it was revealed that the Memphis Police Department created a fake account under the name "Bob Smith" in order to track local activists in the Black Lives Matter movement.[40] "Bob Smith" posed as a sympathizer of protests and was able to successfully "friend" numerous local activists, including a group that planned to stage a die-in at the local mayor's home. Many of these activists later found that they had been put on lists barring them from the mayor's property; some additionally were "arrested several times….on charges related to protesting activity…"[41] Such surveillance therefore had an immediate impact on activists' ability to engage in protest and organizing. Facebook also took action against the Memphis Police Department, disabling the "Bob Smith" account as well as six other accounts they found were linked to the Department.[42]

Though police generally put together fictional personas, such as the "Bob Smith" of the Memphis example, there have been a few occasions where officers have explicitly attempted to pose as real people. The Justice Department was forced to pay a settlement in 2015 after a DEA agent arrested Sondra Arquiett on drug charges, then seized her mobile phone and used it to create a fake Facebook page using her identity and personal photographs. The DEA agent then used the Facebook account to communicate with Arquiett's friends and "friend" other wanted fugitives, in an attempt to identify and arrest others members of Arquiett's drug ring. To give the account the

---

[39] *Id.*
[40] Brentin Mock, *Memphis Police Spying on Activists is Worse Than We Thought,* CITYLAB (Jul. 27, 2018), https://www.citylab.com/equity/2018/07/memphis-police-spying-on-activists-is-worse-than-we-thought/566264/.
[41] Jon Schuppe, *Undercover Cops Break Facebook Rules to Track Protestors, Ensnare Criminals,* NBC NEWS (Oct. 5, 2018), https://www.nbcnews.com/news/us-news/undercover-cops-break-facebook-rules-track-protesters-ensnare-criminals-n916796 The government claimed the woman had ""implicitly consented by granting access to the information stored in her cellphone and by consenting to the use of that information to aid in an ongoing criminal investigation." *Id.*
[42] Id.

appearance of legitimacy, the agent would routinely post as Arquiett, including uploading pictures of her with family members and in her underwear.[43] Facebook supported Arquiett, sending the DEA a letter calling its actions "a knowing and serious breach of Facebook's terms and policies" and asking the DEA to "immediately confirm that it has ceased all activities on Facebook that involve the impersonation of others or that otherwise violate our terms and policies."[44] While the DOJ promised to review its social media policy following the outcry over the Arquiett case, it is unclear that any legitimate changes to their policies have arisen as a result.[45]

This appears to be an area where few, if any, police departments explicitly regulate or conduct oversight of how such undercover operations proceed. *The Root* sent FOIA requests to fifty major police departments nationwide to determine how many had policies in place regulating undercover social media presences. Just thirteen responded with an official policy, and most policies were deemed "fairly simple."[46] The Austin Police Department appeared to be the only department with an "explicit requirement that criminal activity must be suspected to justify the creation of an online alias," meaning for the majority of the departments surveilled, there need not even be a nexus to criminal activity to justify the creation of a fake account.[47] Other departments appeared to allow the creation of fake accounts "at the discretion of individual officers," meaning there is no formal

---

[43] Jose Pagliery, *Facebook Tells DEA: Stop Impersonating Users,* CNN (Dec. 29, 2014), https://money.cnn.com/2014/10/20/technology/security/facebook-dea/?iid=EL

[44] Letter from Joe Sullivan, Chief Security Officer, Facebook Inc., to Michele M. Leonhard, Administrator, Drug Enforcement Administration (Oct. 17, 2014), http://i.cdn.turner.com/money/2014/images/10/20/facebook-letter-to-dea.pdf?iid=EL

[45] Sari Horwitz, *Justice Dept. Will Review Practice of Creating Fake Facebook Profiles*, WASH. POST (Oct. 7, 2014), https://www.washingtonpost.com/world/national-security/justicedept-will-review-practice-of-creating-fake-facebook-profiles/2014/10/07/3f9a2fe8-4e57-11e4- aa5e-7153e466a02d_story.html.

[46] The exceptions were New York City, NY and Austin, TX both of which require supervisor permission to create an account, as well as justification from the supervisor explaining why the account is being created and including information on the profile. Kashmir Hill & Anne Branigin, *Very Few Police Departments Have Rules for Undercover Cops on Facebook,* THE ROOT (Oct. 23,2018), https://www.theroot.com/very-few-police-departments-have-rules-for-undercover-c-1829922261

[47] *Id.*

standard or oversight for the number of profiles created, who they friend, how they interact with other users, and how long such monitoring can be continued.[48]

The prevalence of this phenomenon has drawn criticism of social media companies themselves for failing to fully consider how certain features designed to enhance the user experience may actually enable surveillance in the wrong hands. Snap, for instance, debuted their "SnapMaps" feature last year which pinpointed and updated users' locations in real-time on a map visible to their "friends" through the app. Given that a significant pool of Snap's user base constitutes pre-teens and teenagers, a number of civil liberties and child safety groups pointed out that such detailed location information could be used to paint an accurate picture of one's daily comings and goings, with potentially dangerous ends. [49] In the surveillance context, though there haven't been any reported instances of law enforcement creating fraudulent Snap accounts in this manner, it is easy to envision a similarly dangerous hypothetical. Given that departments like the NYPD have previously used fake accounts largely to monitor the activities of young teenagers, it is not unreasonable to imagine a scenario where law enforcement might create a fake Snap account, "friend" teenagers of interest, and then use Snap Maps to essentially monitor their location in real-time. Snap lacks any clear prohibitions on either creating fraudulent accounts or impersonating another person; or on using its platform for surveillance,[50] thereby placing the onus on teenagers themselves to be hyper-vigilant about interacting with other users on its platform. This is one instance of how social media companies' failure to fully consider the negative ramifications of a product feature might ultimately enable covert surveillance of their users.

---

[48] *Id.*

[49] Alan Loughname, *Snapchat's New Feature Is Not Only Creepy, It's Dangerous,* JOE (2018), https://www.joe.ie/tech/Snaps-feature-creepy-dangerous-593897

[50] *See* appendix A.

C. Formal Legal Requests

Law enforcement has also grown increasingly reliant on assistance from social media platforms and other tech companies to aid their investigations. A CSIS report found that in 2017, U.S. law enforcement made over 130,000 requests for digital evidence to just six tech companies—Google, Facebook, Microsoft, Twitter, Oath (formerly Yahoo!), and Apple—with Facebook and Google getting the bulk of requests.[51] These requests spanned everything from private communications content to metadata (such as location information) and names and IP addresses of particular users.[52] Transparency reports from Twitter and Facebook make clear that there has been a sharp increase, both in the number of requests from law enforcement, and the number of individual accounts targeted. The CSIS report notes that providers reportedly "described deep-seated frustration with what they viewed as overbroad and boilerplate requests from law enforcement…. [i]n particular, providers complained that they were often issued broad-based requests for data that were not, in their view, appropriately tailored."[53] Twitter's transparency report notes, for instance, that in the first half of 2018, they rejected 46% of requests as overbroad, and either denied the information in full, or only responded after the scope of the request was narrowed.[54] Nevertheless, tech companies overwhelmingly acquiesce to such requests for information, only denying around 20% of requests.[55]

Where such requests are granted, they clearly provide law enforcement with extremely detailed content about individual users, from the content of private messages to detailed location information. At the same time, this appears to be the form of surveillance with the most overall regulation. The majority of providers surveyed by this paper will explicitly only comply with legal

[51] Carter & Daskal, *supra* note 9, at 17.
[52] Id.
[53] *Id.* at 19
[54] *Information Requests,* Twitter Inc., https://transparency.twitter.com/en/information-requests.html
[55] Carter & Daskal, *supra* note 9, at 17.

requests when such requests comply with applicable law, including the Stored Communications Act. They also generally require a warrant or court order to release such information.[56] There are accordingly two levels of oversight: first, from the judicial system, and second, from the companies themselves. The high frequency at which warrants are granted,[57] and the relatively high acquiescence from the platforms themselves suggest that such oversight may often be a rubber stamp more than meaningful review—nevertheless, as demonstrated, this is far more review than most other forms of social media surveillance receive.

D. Data Mining and Analytical Software

In recent years, law enforcement agencies have benefitted from a growing market for analytical tools that have the capacity to "mine" social media information, automatically monitor user activity, track users by location or keywords, and conduct other forms of analysis. A study by the Brennan Center found that police departments, cities, and counties nationwide had spent close to $6 million on social media monitoring software, with law enforcement agencies spending almost $5 million of the total.[58]

Such software can perform various functions. Platforms, including Twitter, Facebook, and Instagram, have, in the past, made a variety of user information available to third parties through specialized feeds, including user locations and details of publicly-posted content. Though such access was initially intended for "media companies and brand purposes," in recent years, a number of third-party developers have been able to purchase access to these platforms' API, or application programming interface, a back-end developer tool which allows them to query social media data in

---

[56] *See* appendix A.
[57] *E.g. Data: Most Utah Warrants Approved in Less Than Three Minutes,* ASSOC. PRESS (Jul. 15, 2018). https://www.apnews.com/a2b48c6f1911472986b0e501bdca9f25
[58] Map: Social Media Monitoring by Police Departments, Cities, and Counties, BRENNAN CTR. FOR JUSTICE (Nov. 16, 2016), https://www.brennancenter.org/analysis/map-social-media-monitoring-police-departments-cities-and-counties.

real-time and thus feed data-mining and analytical programs. One such developer, Geofeedia, created a social media monitoring product that was explicitly marketed to, and used by law enforcement as a tool to monitor activists and protestors—its marketing materials referred to unions and activist groups as "overt threats."[59] Geofeedia allows the police to "search Twitter, Facebook, Instagram, Picasa, Flickr and Weibo for keywords in real-time, geographically locating people as they communicate with each other on the go, reading their posts, viewing their pictures and videos, and tracking who they interact with."[60] It also claimed to be able to measure "sentiment" and predict an eruption of violence at protests in Ferguson or Baltimore by analyzing social media posts.[61] Finally, Geofeedia reportedly told potential law enforcement customers that it was finding means of accessing "private Facebook posts" by enabling police to tie in fake or undercover accounts : in order to "track persons of interest across all of their social media sites (Facebook, Twitter, Instagram, etc.) whether their posts are geo-tagged or not."[62]

Examples of police using Geofeedia to surveil protests movements, activists, and organizers nationwide are numerous—often in some of the country's most liberal cities. The ACLU of Northern California revealed that police in Oakland acquired Geofeedia without alerting the city's Privacy Commission (in violation of a municipal ordinance), then reportedly used it to surveil Black

---

[59] Nicole Ozer, *Police Use of Social Media Surveillance Software is Escalating, and Activists are In the Digital Crosshairs,* ACLU (Sept. 22, 2016), https://www.aclu.org/blog/privacy-technology/surveillance-technologies/police-use-social-media-surveillance-software?redirect=blog/free-future/police-use-social-media-surveillance-software-escalating-and-activists-are-digital.

[60] Ali Winston, *Oakland Police Quietly Acquired Social Media Surveillance Tool,* EAST BAY EXPRESS (Apr. 13, 2016), https://www.eastbayexpress.com/oakland/oakland-cops-quietly-acquired-social-media-surveillance-tool/Content?oid=4747526.

[61] Ansel Herz, *How the Seattle Police Secretly—and Illegally—Purchased a Tool for Tracking Your Social Media Posts,* THE STRANGER (Sept. 28, 2016), https://www.thestranger.com/news/2016/09/28/24585899/how-the-seattle-police-secretlyand-illegallypurchased-a-tool-for-tracking-your-social-media-posts This method involved "correlating a positive and negative point score with certain phrases. For example, a negative score for "I'm gonna set off a bomb," versus a neutral score for "I photobombed my friend." *Id.*

[62] Dell Cameron, *Dozens of Police-Spying Tools Remain Available after Facebook, Twitter Crack Down on Geofeedia,* DAILY DOT (Oct. 11, 2016), https://www.dailydot.com/layer8/geofeedia-twitter-facebook-instagram-social-media-surveillance/ At least one other similar company claimed to be able to "defeat user efforts to conceal their locations on Twitter" because of "exclusive access to Twitter's back end." *Id.*

Lives Matter protests in the city.[63] Police in Seattle similarly acquired Geofeedia without approval from the local City Council, in violation of a municipal privacy ordinance.[64] The ACLU also found that according to Geofeedia's own promotional material, amid protests following the killing of Freddie Gray by Baltimore police, the same police department used Geofeedia "to run social media photos through facial recognition technology" to identify protestors "with outstanding warrants and arrest them directly from the crowd."[65]

Geofeedia enabled surveillance of other protests and movements as well. Media reports indicated that the San Jose police utilized Geofeedia to surveil South Asian protestors during the visit of the Indian prime minister Narendra Modi. [66] The ACLU of Massachusetts released a report finding that the Boston Police Department used Geofeedia to disproportionately target the Muslim community.[67] Documents revealed that the BPD searched for keywords they identified as "Islamic extremist terminology" including words like "ISIS" and "caliphate."[68] They also tracked the hashtag #muslimlivesmatter, as well as common Arabic words like "ummah," which means "community."[69]

The Geofeedia example demonstrates the real role that platforms can play in safeguarding their data and tools from feeding into surveillance programs. Within days of ACLU's revelations and the resulting media scrutiny, Facebook, Instagram and Twitter cut off Geofeedia's access to their APIs, and took steps to ensure that similar companies would also be restricted from gaining access

---

[63] Winston, *Oakland Police Quietly Acquired Social Media Surveillance Tool, supra* note 61.
[64] Herz, *How the Seattle Police Secretly—and Illegally—Purchased a Tool, supra* note 62.
[65] Press Release, Geofeedia, *Baltimore County Police Department and Geofeedia Partner to Protect the Public During Freddie Gray Riots* (undated), https://www.aclunc.org/docs/20161011_geofeedia_baltimore_case_study.pdf
[66] Winston, *Oakland Police Quietly Acquired Social Media Surveillance Tool, supra* note 61.
[67] Alanna Durkin Richer, *Boston Police's Social Media Surveillance Unfairly Targeted Muslims, ACLU Says,* Boston Globe (Feb. 7, 2018), https://www.bostonglobe.com/metro/2018/02/07/boston-police-social-media-surveillance-unfairly-targeted-muslims-aclu-says/9JUpzPmy8Tsr5RLxvCm61M/story.html
[68] *Id.*
[69] *Id.*

to this data.[70] All three entities further released statements clarifying that their policies do not permit developers to use their data for surveillance.[71]

Despite the crackdown on Geofeedia, the market is awash with similar—or even more intrusive—programs. A number of companies have actively propagated algorithmic predictive policing applications that factor in social media data to essentially predict both geographic locations where crime might surge; as well as individuals' propensity to either commit crimes or become victims. The ACLU found that the Fresno Police Department had been using a software called Beware, which "gathers information on a person's publicly available social media activity, and assigns them a threat level of green, yellow, or red"—one of the people labeled at threat level yellow included a local member of the Fresno City Council who had been outspoken on civil rights and racial justice issues.[72] Fresno police also used a software program called "MediaSonar" whose promotional material "encouraged the police to identify "threats to public safety" by tracking #BlackLivesMatter-related hashtags, including #dontshoot and #imunarmed." MediaSonar's materials suggested that these "keywords" could "help identify illegal activity and threats to public safety."[73] Because these programs track publicly available social media postings rather than third-party developer tools, they appear to still be functioning.

Similarly, Hitachi's Visualization Predictive Crime Analytics program for feeds in social media data, such as public tweets, and extrapolates information from geotagged tweets to find

---

[70] Chris Moody, *Developer Policies to Protect People's Voices on Twitter,* Twitter (Nov. 22, 2016), https://blog.twitter.com/developer/en_us/topics/community/2016/developer-policies-to-protect-peoples-voices-on-twitter.html

[71] *Id.* ("Using Twitter's Public APIs or data products to track or profile protesters and activists is absolutely unacceptable and prohibited. We prohibit developers using the Public APIs and Gnip data products from allowing law enforcement — or any other entity — to use Twitter data for surveillance purposes. Period.")

[72] Matt Cafle, *This Surveillance Software Is Probably Spying on #BlackLivesMatter,* ACLU (Dec. 15, 2015), https://www.aclunc.org/blog/surveillance-software-probably-spying-blacklivesmatter.

[73] *Id.*

common terms and link them to increased crime.[74] The Boston Police Department used an IBM facial recognition software system to surveil attendees of a major concert; the software captured faces of thousands of concert attendees, then "filtered their appearance into data points which could then be cross-checked against certain identifying characteristics."[75] The data then fed into a hub "where city representatives, Boston Police, and IBM support staff could watch in real time, all while simultaneously monitoring social media key words related to the event."[76] Palantir similarly patented and provided a crime-forecasting software to numerous police departments nationwide that incorporates social media analyses as one tool in a larger algorithm intended to "predict the likelihood that individuals would commit violence or become victims."[77] For instance, in New Orleans, Palantir provided software to the New Orleans Police Department that "traced people's ties to other gang members, outlined criminal histories, analyzed social media, and predicted the likelihood that individuals would commit violence or become a victim." As in the Geofeedia examples, the NOPD's acquisition of Palantir software escaped municipal scrutiny because it was billed by Palantir as a "philanthropic relationship with the city" rather than a surveillance tool.[78]

These surveillance techniques are not limited to the United States. Israeli security forces have made considerable investments in predictive intelligence systems that monitor and analyze online activities of Palestinians, with the goal of extrapolating information to prevent terror attacks.[79] In 2018, Israeli authorities claimed they had foiled over 200 terror attacks through such monitoring.[80]

---

[74] Jack Smith IV, *Police Are Sweeping Up Tweets and Friending You on Facebook, Whether You Know It or Not,* MIC (Nov. 11, 2015), https://mic.com/articles/128299/how-police-use-twitter-and-facebook-to-predict-crime#.mtM2C38la
[75] Luke O'Neil, *Beantown's Big Brother: How Police Used Facial Recognition Technology to Spy on Thousands of Music Festival Attendees,* VICE (Aug. 13, 2014), https://noisey.vice.com/en_us/article/6wm356/beantowns-big-brother
[76] *Id.*
[77] Ali Winston, *Palantir Has Secretly Been Using New Orleans To Test Its Predictive Policing Technology,* THE VERGE (Feb. 27, 2018), https://www.theverge.com/2018/2/27/17054740/palantir-predictive-policing-tool-new-orleans-nopd
[78] Key members of the city council were not aware of Palantir's work in New Orleans until approached by media. *Id.*
[79] *Police Minister: Social Media Monitoring Has Foiled 200 Terror Attacks,* Times of Israel (Jun. 12, 2018), https://www.timesofisrael.com/police-minister-social-media-monitoring-has-foiled-200-terror-attacks/; Orr Hirschauge & Hagar Shezaf, *How Israel Jails Palestinians Because They Fit The "Terrorist Profile,"* HAARETZ (May 31, 2017), https://www.haaretz.com/israel-news/.premium-1.792206.
[80] *Id.*

Algorithmic software exponentially increases the privacy and constitutional concerns raised by the other forms of social media surveillance discussed in this paper. Where undercover accounts or manual surveillance of publicly posted materials still face some resource constraints, algorithmic software can capture and analyze troves of data that it would take humans centuries to analyze similarly. Many critics have also raised serious concerns that such programs often serve to entrench and reinforce existing biases in the criminal justice system, while often generating inaccurate conclusions about an individual's future propensity to commit crimes. [81] The secretive and proprietary nature of the algorithms themselves make it almost impossible for outside observers to accurately evaluate the level of inherent bias such programs may contain. Despite the significant cost of these programs, they may do little to actually accurately predict crime or future criminality; and once again, disproportionately impact communities of color.

In this manner, even as activists have grown increasingly reliant on social media platforms to spread their message, organize, and further their movements, law enforcement agencies are equally reliant on such platforms, expanding beyond ordinary crime-fighting purposes to broader surveillance measures of these movements. Even in instances where police are not explicitly surveilling political or social movements, the above discussion indicates that social media surveillance is likely to disproportionately impact people of color, as the example of the NYPD's anti-gang efforts indicate. This sort of widespread surveillance in turn results in a marked chilling effect on such communities. A 2016 study found that for the majority of respondents, being aware

[81] For instance, a ProPublica investigation found that an algorithm used to create "risk scores" for recently arrested individuals in order to predict the likelihood of recidivism and set bail amounts was "remarkably unreliable in forecasting violent crime: [o]nly 20 percent of the people predicted to commit violent crimes actually went on to do so." Julia Angwin et al., *Machine Bias: Risk Assessments in Criminal Sentencing,* ProPublica (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing. Similarly, a 2016 study of the software PredPol, which uses historical crime data to predict future crime activity, found that it "replicated "systemic bias" against over-policed communities of color" and historical crime data generated inaccurate predictions about future crime. Kristian Lum & William Isaac, *To Predict And Serve?* Significance Mag. (Oct. 2016), https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1740-9713.2016.00960.x.

of government surveillance "significantly reduced the likelihood of speaking out in hostile opinion climates."[82] The study noted that the perception of being surveilled tended to produce significantly conformist behaviors, noting that when individuals "perceive they are being monitored, they readily conform their behavior—expressing opinion when they are in the majority and suppressing them when they are not."[83] As such, even when police engage in such surveillance with the goal of fighting crime, they may unintentionally chill legitimate speech and protest when such efforts are disproportionately directed at a particular community. This is exemplified, for instance, by the NYPD's secret surveillance of the New York Muslim community following 9/11. A comprehensive report found that the "chilling effect" of such surveillance had been marked and severe. From congregants refusing to attend religious services, to students growing afraid to discuss politics and civil rights, or engage in activism at their educational institutions, the awareness that the Muslim community as a whole was under surveillance chilled and constrained thousands of law-abiding Muslims from fully participating in religious, political, and communal spaces.[84] Such surveillance therefore, has clear implications for the First, Fourth, and Fourteenth Amendment Rights of affected communities of color. The impact of social media surveillance on constitutional rights will be explored further in the next section.

PART II: CONSTITUTIONAL IMPLICATIONS OF SOCIAL MEDIA SURVEILLANCE

A full analysis of the constitutional rights implicated by such surveillance is beyond the scope of this paper.[85] In brief, however, such surveillance raises concerns about the potential for

---

[82] Elizabeth Stoychff, *Under Surveillance: Examining Facebook's Spiral of Silence Effects in the Wake of NSA Internet Monitoring,* 93 Journalism & Mass Comm. Quarterly 296 (2016). *See also* Nafeez Ahmed, *'Chilling Effect' of Mass Surveillance Silencing Dissent Online, Study Says,* Vice (Mar. 17, 2016), https://motherboard.vice.com/en_us/article/aekedb/chilling-effect-of-mass-surveillance-is-silencing-dissent-online-study-says

[83] Ahmed, *'Chilling Effect' of Mass Surveillance, supra* note 84.

[84] *See generally* Diala Shamas & Nermeen Arastu, Muslim American Civil Liberties Coalition, Mapping Muslims (2014), http://www.law.cuny.edu/academics/clinics/immigration/clear/Mapping-Muslims.pdf.

[85] A number of other scholars have examined these constitutional implications in greater detail. *See, e.g.,* Rachel Levinson Waldman, *Government Access to and Manipulation of Social Media: Legal and Policy Challenges,* 61 Howard L.J. 523 (2018);

infringement of individuals' First, Fourth, and Fourteenth Amendment rights. The Fourth Amendment, which protects individuals' rights to be free from unreasonable searches and seizures, has also been interpreted to protect individuals from government surveillance in public spaces where individuals have a "reasonable expectation of privacy" in the information collected.[86] Of course, one can hardly argue that people have a reasonable expectation of privacy in information they publicly and voluntarily post to social media. Similarly, in a long string of cases, courts have generally held that undercover law enforcement activity does not violate individuals' Fourth Amendment rights.[87] There have been several where courts have explicitly found that the use of undercover personas on social media platforms—either by law enforcement directly, or through the use of informants—does not violate the Fourth Amendment, even where users had set up privacy controls (such as a private forum or profile) to evade scrutiny.[88] When individuals voluntarily post content or use social media platforms, therefore, even the use of platform privacy controls does not suggest that users have maintained a "reasonable expectation of privacy" in the material posted.

Yet, a growing majority on the Supreme Court has acknowledged that in the context of rapidly advancing technology, law enforcement's capacity to conduct surveillance requires far less money and manpower and can be accomplished with greater ease than ever before—and accordingly, there may be a need to rethink the Fourth Amendment in the digital age. In a number

---

[86] Katz v. United States, 389 U.S. 347 (1967)

[87] *See, e.g.,* Hoffa v. United States, 385 U.S. 293 (1966) (finding that evidence derived from defendant's conversations with an undercover government informant was not the product of an unlawful Fourth Amendment search, holding that the "Fourth Amendment [does not] protect[] a wrongdoer's misplaced belief that a person to whom he voluntarily confides his wrongdoing will not reveal it."); Guest v. Leis, 255 F.3d 325 (6th Cir. 2001)

[88] Guest v. Leis, 255 F.3d 325 (6th Cir. 2001) (finding that an officer's use of an undercover account to access an online bulletin board system did not violate the Fourth Amendment because its users had no legitimate expectation of privacy in information they had voluntarily disclosed to the bulletin board's system operator.); US v. Meregildo 883 F. Supp. 2d 523 (S.D.N.Y. 2012) (finding that police's creation of a fake Instagram account to view posts in private account did not violate the Fourth Amendment because the user's legitimate expectation of privacy ended when he disseminated content to his friends); Palmieri v. United States, 72 F. Supp. 3d 191, 210 (D.D.C. 2014) (the account holder had no reasonable expectation of privacy in data shared voluntarily with a "friend," even if that friend turned out to be a government agent).

of recent cases, the Supreme Court has sided with criminal defendants in cases where the government's evidence was sourced through the collection of vast amounts of digital data—even when such data was technically in the public sphere.[89] Under the "mosaic theory" advanced by Justice Sotomayor in the seminal case *United States v. Jones,* while individuals may not have a compelling reasonable expectation of privacy over individual pieces of data, when such data is aggregated by law enforcement in a manner that "generates a precise, comprehensive record of a person's public movements that reflects a wealth of detail about her familial, political, professional, religious, and sexual associations," their privacy interests are exponentially higher and accordingly, worthy of constitutional protection.[90] By analogy, it seems reasonable to extend this to the context of social media. Even where people might make individual social media posts publicly available, it may seem plausible to argue that they have a reasonable expectation of privacy over the whole of their social media activity over an extended period of time, particularly when analytical tools might be applied to generate the "precise, comprehensive" record of their movements and activities described by Justice Sotomayor in *Jones*. Few people anticipate that someone is tracking or aggregating each piece of content they post over a period of years—whether this is done by individual officers, as with the NYPD's gang work—or in a more automated fashion, through software like Geofeedia or Palantir's crime prediction platform. Though the Court has not considered this question in the context of social media surveillance (and there are certainly far more complexities to this analogy than this brief analysis can provide) its general shift towards a stronger embrace of updating the Fourth Amendment's protections for the digital age is particularly

[89] For instance, in *Riley v. California,* the Court unanimously held that law enforcement must obtain a warrant before searching an individual's cellphone finding that "[m]odern cellphones… implicate privacy concerns far beyond those implicated by [physical searches]. 134 S. Ct. 2473, 2488–89 (2014); similarly in *Carpenter v. United States,* the Court similarly held that the warrantless search and seizure of cell-site records violated the Fourth Amendment, acknowledging that expectations of privacy in the digital era have evolved beyond existing precedents. 138 S. Ct. 2206 (2018).
[90] United States v. Jones, 132 S. Ct. 945, 955 (2012) (Sotomayor, J., concurring).

promising in considering constraints on algorithmic or predictive software that uses social media in its analyses.

Similarly, in the context of the First Amendment, at least one circuit court has affirmatively held that both "likes" and comments on Facebook constitute First Amendment-protected speech,[91] while the Supreme Court has found that government policies "foreclose[ing] access to social media altogether… prevent the user from engaging in the legitimate expression of First Amendment rights."[92] There is therefore growing recognition that activity on social media platforms can constitute protected speech. The chilling effect engendered by social media surveillance may therefore provide a basis for a First Amendment challenge to such activities. This is likely to be an uphill battle: the Supreme Court explicitly held in *Laird v. Tatum* that mere "allegations of a subjective chill" fostered by a "governmental investigative and data-gathering activity" is insufficient to confer standing on individuals to challenge such activity without a showing of a "specific present objective harm or a threat of specific future harm."[93] However, a recent Third Circuit case suggests that where such a surveillance program is applied discriminatorily, and accordingly dissuades individuals from a particular protected class from exercising their constitutional rights, it may well violate the First Amendment. In *Hassan v. City of New York,* which challenged the NYPD's post-9/11 surveillance of New York Muslims, the Third Circuit distinguished *Laird* on grounds that the challenged surveillance there emanated from a military data-gathering program that had developed in response to civil rights protests, but nevertheless appeared to be applied in a non-discriminatory matter.[94] In contrast, the court found that the challenged surveillance in *Hassan* intentionally targeted

[91] Bland v. Roberts, 730 F.3d 368, 385, 388 (4th Cir. 2013).
[92] Packingham v. North Carolina, 137 S. Ct. 1730, 1735 (quoting Reno v. American Civil Liberties Union, 521 U. S. 844, 868 (1997)).
[93] Laird v. Tatum, 408 U.S. 1, 2, 13-14 (1972) (holding that "the mere existence, without more, of a governmental investigative and data-gathering activity that is alleged to be broader in scope than is reasonably necessary for the accomplishment of a valid governmental purpose" is not violative individual's First Amendment rights).

[94] 804 F.3d 277 (3d Cir. 2015)

Muslims, often using ethnicity as a proxy for religion. Moreover, unlike the *Laird* plaintiffs, who had failed to articulate a specific injury, the court found that because the NYPD's surveillance program was "carried out in a discriminatory manner" and caused "direct, ongoing, and immediate harm" to the plaintiffs' ability to engage in their religious and other First-Amendment-related activities.[95] The court made clear that "*Laird* doesn't stand for the proposition that public surveillance is… *per se* immune from constitutional attack," particularly when such discriminatory attributes and clear harms can be demonstrated.[96] The *Hassan* court also found the NYPD's surveillance program to be constitutionally suspect under the Fourteenth Amendment, finding that the plaintiffs had plausibly alleged that the NYPD's program was facially discriminatory, and had in fact monitored Muslims "with a greater degree of severity than other… groups."[97] It also rejected the city's assertion that the program was permissible because it was motivated by legitimate law enforcement purposes of protecting the public safety, and any discrimination that may have resulted was not the result of "ill will, enmity, or hostility."[98] The limited conclusion *Hassan* offers in the context of social media surveillance indicates that such surveillance may implicate the First and Fourteenth Amendments where such surveillance is wielded against a particular group "with a greater degree of severity than other… groups," even if such surveillance is justified by legitimate law enforcement purposes.

The law, particularly where digital privacy is concerned however, is generally slow to evolve. And despite promising precedents like *Hassan,* few courts have explicitly considered the constitutionality of social media surveillance techniques and it is far from certain that challenges to such programs would be successful in court. However, in instances where the legal system has been slow to protect digital privacy, social media companies themselves can play a significant role in

[95] *Id.* at 292
[96] *Id.*
[97] *Id.* at 298.
[98] *Id.* at 297.

ensuring that their services and data are not being co-opted for surveillance. Take Geofeedia's example. Media and civil rights organizations demonstrated that law enforcement were using Geofeedia to surveil individuals engaged in First Amendment-protected protests and organizing activities. Such surveillance didn't end because of a court order—rather, it came after Facebook, Twitter, and Instagram suspended Geofeedia's access to their data and back-end developer tools.[99] In another instance, the ACLU revealed that a company called Dataminr was marketing a product to law enforcement that would allow them to conduct location-based tracking and search "billions of real-time and historical tweets" based on its access to Twitter's API.[100] Dataminr advertised its products as a means of tracking and surveilling protests to its clients, including the Los Angeles police department; and provided its software to all 77 federally funded "fusion centers" nationwide.[101] Twitter responded by requiring Dataminr to terminate its contracts with all fusion center accounts and publicly commit not to provide government customers with data access or features that allow for "any form of surveillance."[102]

Companies therefore can play a key role by choosing to either facilitate law enforcement's surveillance of communities of color—or refusing to cooperate. A good starting point here is the platforms' own internal guidelines and policies regarding user privacy, data retention, and law enforcement use. These policies provide a general understanding of how law enforcement is permitted to use or request data from social media providers; and how platforms themselves regulate such uses. The next section will review and compare the relevant guidelines from four

[99] *See supra* notes 60-72 and accompanying text.
[100] Nicole A. Ozer, *Twitter Cuts Off Fusion Spy Centers' Access to Social Media Surveillance Tool,* ACLU (Dec. 15, 2016), https://www.aclunc.org/blog/twitter-cuts-fusion-spy-centers-access-social-media-surveillance-tool
[101] *Id.*
[102] *Id.* Twitter's letter notes that a Dataminr news alert service will still be available to law enforcement and organizations supporting first responders.

major social media platforms to determine the extent to which such companies facilitate law enforcement's use of their data and platform for surveillance.

PART III: ASSESSING PLATFORMS' INTERNAL PRIVACY POLICIES AND LAW ENFORCEMENT GUIDELINES

This analysis relied on publicly-available policies and guidelines from the four companies surveyed—Facebook, Instagram, Twitter, and Snap (the owner of Snap). These companies were chosen both for their size—they represent some of the largest social media platforms and user bases in the field today—and because each site's ethos rests largely on social networking and communications through different mediums. Though companies like Apple and Google may also play roles in law enforcement surveillance of particular communities, they have not been included in this analysis because their business model does not rely on social networking and communications to the same extent. These four companies therefore provide stronger comparative value because the products they offer bear more substantial similarities to each other.

The fact that such policies are even publicly available to begin with suggests a significant shift. In 2011, EFF conducted a study "investigating how the government seeks information from social networking sites such as Twitter and how the sites respond to these requests."[103] This required access to many of the same policies this analysis used, but EFF was then forced to submit a federal Freedom of Information Act request to the Department of Justice to obtain each company's set of law enforcement guidelines.[104]

Briefly, this analysis considers: 1) what information is available on each website, including the visibility of user profiles to individuals outside the network, user privacy controls, and encryption

[103] Jennifer Lynch, *Social Media and Law Enforcement: Who Gets What Data and When?* Elec. Frontier Found. (Jan. 20, 2011), https://www.eff.org/deeplinks/2011/01/social-media-and-law-enforcement-who-gets-what.
[104] *Id.*

options; 2) the types of information collected and retained by the platforms themselves and to what extent this information is available to law enforcement, including information on the availability of deleted content; 3) the level of legal process required for law enforcement to formally obtain user data, including whether platforms notify users when their data is sought; 4) whether the platforms explicitly ban the use of their data for surveillance purposes, and how robust such definitions and prohibitions are; and 5) to what extent platforms' transparency reports account for government surveillance. A table detailing the findings in each category is available at Appendix A.

i. **Non-User Visibility and User Privacy Controls**

Twitter and Instagram do not require individuals to create a user account before being able to view other users' profiles or posted content. Both provide users with the option to set their accounts to "private," which restricts viewing only to approved followers, but users generally cannot control privacy settings beyond that for individual pieces of content.[105] Facebook will show non-users only limited information about users, and unlike Twitter and Instagram, allows users to modulate privacy settings for each piece of content posted, allowing individuals to hide specific posts or photographs even from accepted Facebook "friends."[106] Users who are not "friends" with another account-holder will similarly see only limited information about that user. Snap requires users to have a Snap account in order to view others' stories. By default, only "friends" added by the user can either directly contact, view the user's posted Stories, or see the user's location on SnapMaps.[107] Snap does allow users additional privacy controls, including the ability to hide stories from specific users.

---

[105] The exception is Instagram Stories' recently launched "Close Friends" option, which restricts the viewing of a story to a pre-determined "close friends" list. *See* Press Release, Instagram *Share With Your Close Friends On Instagram Stories* (Nov. 30, 2018), https://instagram-press.com/blog/2018/11/30/share-with-your-close-friends-on-instagram-stories/.
[106] *How Can I Adjust My Privacy Settings?* Facebook Help Center, https://www.facebook.com/help/193677450678703?helpref=uf_permalink (last visited Jan. 21, 2019).
[107] Snap does have a public option, allowing all Snap users to contact an individual. *Privacy Settings,* Snapchat Support, https://support.snapchat.com/en-US/a/privacy-settings2 (last visited Jan. 21, 2019).

ii. **Policies and Reporting**

All platforms surveyed provided easily accessible links to their guidance for law enforcement, data retention policies, and privacy policies.[108] It is worth noting that users may need to jump around from various sets of guidelines to gain a full accounting of how these policies interact—law enforcement policies, for instance, contain little information about categories of information stored and data retention policies. Facebook, Twitter, and Snap also provide Transparency Reports with varying levels of detail.[109] Instagram does not appear to publish a separate Transparency Report, and its parent company Facebook does not disaggregate its own Transparency Report to indicate the number of legal requests received by Instagram, rather than the main Facebook entity.[110]

In general, the transparency reports provide information on the number of legal requests received by each platform from governmental entities, and the overall rate of compliance with such requests. Across the board, compliance with law enforcement requests appears to be extremely high, with all four platforms averaging between 76-88% compliance with requests from US governmental agencies.[111] Twitter's report also notes that it pushes back on requests that are overbroad or improper, and that it narrowed or did not disclose information in 46% of cases.[112] No platform appears to provide a breakdown of the requesting governmental entities, for instance by delineating non-law enforcement requestors from law enforcement requestors.

iii. **Information Available to Law Enforcement and Required Legal Processes:**

---

[108] *See, e.g., Information for Law Enforcement,* Instagram, https://help.instagram.com/494561080557017 (last visited Jan. 21, 2019); *Information for Law Enforcement Authorities,* Facebook, https://www.facebook.com/safety/groups/law/guidelines/ (last visited Jan. 21, 2019); *Information for Law Enforcement,* Snap, https://www.snap.com/en-US/safety/safety-enforcement/ (last visited Jan. 21, 2019); *Guidelines for Law Enforcement,* Twitter (last visited Jan. 21, 2019), https://help.twitter.com/en/rules-and-policies/twitter-law-enforcement-support

[109] E.g. *Facebook Transparency Report,* Facebook, https://transparency.facebook.com/ (last visited Jan. 21, 2019); *Twitter Transparency Report,* Twitter, https://transparency.twitter.com/en.html (last visited Jan. 21, 2019); *Transparency Report,* Snap Inc., https://www.snap.com/en-US/privacy/transparency/ (last visited Jan. 21, 2019).

[110] *Facebook Transparency Report,* Facebook, https://transparency.facebook.com/ (last visited Jan. 21, 2019); *Twitter Transparency Report,* Twitter, https://transparency.twitter.com/en.html (last visited Jan. 21, 2019); *Transparency Report,* Snap Inc., https://www.snap.com/en-US/privacy/transparency/ (last visited Jan. 21, 2019).

[111] *Twitter Transparency Report,* Twitter, https://transparency.twitter.com/en.html (last visited Jan. 21, 2019).

[112] *Id.*

The platforms all acknowledge collecting, storing, and using similar information: basic subscriber information on users, such as user names, length of service, associated email addresses and phone numbers, credit card information, and IP addresses from recent login attempts; information about content uploaded to the platform, including embedded metadata, such as the location of a photo or the date a file was created on Instagram; information about associated groups and users, "including the types of content you view or engage with; the features you use; the actions you take; the people or accounts you interact with; and the time, frequency and duration of your activities"[113]; and location data. On Twitter, for instance, if users enable "Precise Location," Twitter can collect, store, and use your exact longitude and latitude; and this information will be associated with your Tweet and findable via API.[114]

All four platforms will disclose records solely in accordance with their platforms' terms of service and pursuant to applicable law, including the Federal Stored Communications' Act. In general, all four platforms draw a distinction between "basic subscriber information" and content-based information. Basic subscriber information, such as user names, length of service, associated email addresses and phone numbers, credit card information, and IP addresses from recent login attempts, can be disclosed pursuant to a valid warrant, subpoena, or court order; while disclosure of stored contents of accounts can only be facilitated pursuant to a valid order. This includes communications, including private messages, photographs and videos, and location information. Facebook, Instagram and Snap also note disclosing IP logs, though Instagram and Facebook emphasize that "IP logs are limited and frequently incomplete," [115] while Twitter suggests that given

[113] See Appendix A.
[114] *Tweet Location FAQs,* Twitter, https://help.twitter.com/en/safety-and-security/tweet-location-settings (last visited Jan. 21, 2019).
[115] See Appendix A.

its "real-time nature," information like IP logs may only be stored for a very brief period of time.[116] All four sites are somewhat vague on whether all the information they collect, such as your interactions with other users on the platform, or the "time, duration, and frequency of your activities"—information that is largely collected to provide enhanced information on platform usage—can also be turned over to law enforcement.[117]

iv. **Retention and Availability of Deleted Content**

Twitter, Instagram and Facebook are all extremely vague on their retention policies, with Twitter noting they "retain different types of information for different time periods."[118] Facebook and Instagram state they "store data until it is no longer necessary to provide our services and Facebook Products, or until your account is deleted - whichever comes first" and provide a few examples of the "case-by-case determination" this involves—for instance, that logs of search records are stored for six months, while copies of government-issued ID collected for account verification purposes are deleted after thirty days.[119] Snap provides a more comprehensive assessment of their policies, noting that their servers "automatically delete a Snap after being viewed by intended recipients" and will generally delete unopened Snaps and messages between 24 hours and thirty days. It is somewhat vaguer about Snaps sent to the crowdsourced "Our Story," which is publicly viewable, noting merely that Our Story snaps "may be saved for longer periods of time."[120] It similarly notes varying retention periods for location data.[121]

---

[116] *Guidelines for Law Enforcement,* Twitter (last visited Jan. 21, 2019), https://help.twitter.com/en/rules-and-policies/twitter-law-enforcement-support
[117] *Data Policy,* Instagram, https://help.instagram.com/155833707900388.
[118] *Id.; Facebook Data Policy,* Facebook, https://www.facebook.com/full_data_use_policy .
[119] *Information for Law Enforcement,* Snap, https://www.snap.com/en-US/safety/safety-enforcement/ (last visited Jan. 21, 2019).
[120] *Id.*
[121] *Privacy Policy,* Snap, https://www.snap.com/en-US/privacy/privacy-policy/ ("if you use the Map, we store information about your favorite places for up to 40 days so we can show you Actionmoji and improve your experience.

All four platforms note that generally, if users delete their accounts, "most" user information associated with the accounts is also deleted and cannot be retrieved.[122] No platform specifies what, if any, information is retained when an account is deleted. Twitter and Snap note that when users delete pieces of individual content (such as a Tweet), such information similarly cannot be retrieved.[123] Facebook's policy states that when users delete content, the information is "remove[d] from the site," but while some information is "permanently deleted from our servers" some things "can only be deleted when you permanently delete your account."[124] It does not elaborate on what types of content can be distinguished in this manner. Instagram notes that it "retain[s] information from accounts disabled for terms violations for at least a year to prevent repeat abuse or other term violations."[125] Both Instagram and Facebook note that "[i]nformation that others have shared about you isn't part of your account and won't be deleted."[126]

v. **User Notification**

All four platforms' stated policy is to notify users of requests for their account information, unless prohibited by law or court order from doing so; or in the event of "exceptional circumstances" such as "cases involving child exploitation," cases involving "the threat of imminent death or bodily injury," or in Twitter's case, terrorism.[127] Twitter provides a copy of the request along with the user notification, while Snap provides affected users seven days to challenge the

---

If location information is associated with a Snap—like those saved to Memories or posted to Our Story—we'll retain that location as long as we store the Snap.")

[122] *See* Appendix A.

[123] Snap emphasizes that "because Snap's servers are designed to automatically delete most user content, they "often cannot retrieve user content except in very limited circumstances." *Information for Law Enforcement,* Snap, https://www.snap.com/en-US/safety/safety-enforcement/ (last visited Jan. 21, 2019).

[124] *What Happens To Content (Posts, Pictures) That I Delete From Facebook?* Facebook Help Center, https://www.facebook.com/help/356107851084108.

[125] *Data Policy,* Instagram, https://help.instagram.com/155833707900388.

[126] *Id.*

[127] *See* Appendix A.

request in court before responding.[128] Facebook and Instagram also note that "if [a] data request draws attention to an ongoing violation of our terms of use, [they] will take action to prevent further abuse, including actions that may notify the user that [they] are aware of their misconduct."[129] Snap's policy appears to be relatively new—it only began providing notice in November 2015.[130]

vi. **Policy on Fake/Fraudulent Accounts**

Facebook is the only entity to carry a "real name" policy, requiring users "to use the name they go by in everyday life" and forbidding users from maintaining multiple accounts.[131] It further states "[o]perating fake accounts, pretending to be someone else, or otherwise misrepresenting your authentic identity is not allowed, and we will act on violating accounts."[132] Twitter and Instagram both affirmatively state that they do not require people to use their real names or identities;[133] Twitter explicitly allows the use of pseudonyms and users are allowed to create and manage multiple Twitter accounts. Twitter and Instagram also do not require any form of identity verification.[134] Snap does not provide any guidance on the subject.

vii. **Exigent Circumstances**

All four platforms will consider requests for information in exigent circumstances that involve the danger of death or serious physical injury to a person.[135] Facebook and Instagram will

---

[128] *Guidelines for Law Enforcement,* Twitter (last visited Jan. 21, 2019), https://help.twitter.com/en/rules-and-policies/twitter-law-enforcement-support; *Guidelines for Law Enforcement,* Twitter (last visited Jan. 21, 2019), https://help.twitter.com/en/rules-and-policies/twitter-law-enforcement-support

[129] *., Information for Law Enforcement,* Instagram, https://help.instagram.com/494561080557017 (last visited Jan. 21, 2019); *Information for Law Enforcement Authorities,* Facebook, https://www.facebook.com/safety/groups/law/guidelines/ (last visited Jan. 21, 2019);

[130] *Information for Law Enforcement,* Snap, https://www.snap.com/en-US/safety/safety-enforcement/ (last visited Jan. 21, 2019).

[131] *Misrepresentation—Community Standards,* Facebook Inc., https://www.facebook.com/communitystandards/misrepresentation.

[132] *Id.*

[133] See Appendix A.

[134] *Id.*

[135] *Id.*

additionally consider emergency requests for information in matters "involving imminent harm to a child."[136]

viii. **Encryption**

Snap is the only entity of the four surveilled to offer any sort of encryption protections. Its internal cloud-storage mechanism, "Memories," allows users to store "sent or unsent Snaps, posted Stories, and photos and videos from their phone's photo gallery."[137] Users can enable an encryption option, "My Eyes Only." If enabled, Snap cannot decrypt stored information. [138]

ix. **Use of Platform for Surveillance**

Twitter, Instagram and Facebook explicitly ban third-party developers from using their data to "provide tools that are used for surveillance," but Twitter's definition of what constitutes "surveillance" is substantially more robust and detailed than Facebook and Instagram's definition. Twitter "explicitly prohibits" the use of Twitter content and information derived from it to "any public sector entity" for surveillance purposes, including "investigating or tracking Twitter's users… tracking or monitoring sensitive events (including but not limited to protests, rallies, or community organizing meetings)…analyses or research that isolates… individuals for any unlawful or discriminatory purpose or in a manner that would be inconsistent with… users' reasonable expectations of privacy" and uses by any entity that "target, segment, or profile individuals" based on a host of characteristics, including "racial or ethnic origin, negative financial status or condition, and data relating to any alleged or actual commission of a crime…"[139] In contrast, Facebook and Instagram merely require developers to "protect the information you receive from us against unauthorized access, use, or disclosure. For example, don't use data obtained from us to provide

---

[136] *Id.*
[137] *Information for Law Enforcement,* Snap, https://www.snap.com/en-US/safety/safety-enforcement/ (last visited Jan. 21, 2019).
[138] *Id.*
[139] *Developer Agreement & Policy,* Twitter, https://developer.twitter.com/en/developer-terms/agreement-and-policy.html

tools that are used for surveillance."[140] Snap doesn't appear to have any explicit prohibitions on third party developers using their data for surveillance-related purposes.

Along similar lines, only Facebook appears to directly address law enforcement's use of their platform during investigations, and only in conjunction with their misrepresentation, noting in their law enforcement guide that it will ""always disable accounts that supply false or misleading profile information or attempt to technically or socially circumvent site privacy measures," even to the extent this interferes with law enforcement investigations.[141] Twitter, Instagram and Snap do not provide any guidance or restrictions on law enforcement themselves utilizing the platform for surveillance or during investigations.[142]

PART IV: IMPROVEMENTS AND RECOMMENDATIONS

From the above analysis, perhaps the most glaring gap is that despite the four companies' stated commitments to racial justice and support of movements like Black Lives Matter, all four lack specific prohibitions against using their platforms for surveillance. Though Twitter, Instagram, and Facebook all prohibit third-party developers from building tools or channeling data in ways that facilitate surveillance, there is no explicit restriction on users of the platform itself to refrain from engaging in unlawful surveillance of users' constitutionally protected activities. The most ambitious recommendation here would be for companies to revise their Community Standards or Terms of Use to explicitly prohibit their platform from being used for surveillance or investigations without a clear nexus to criminal activity. This would at least draw a line between using social media to gather evidence of legitimate wrongdoing, versus merely accumulating intelligence on individuals engaged

[140] *Platform Policy,* Facebook for Developers, https://developers.facebook.com/policy/
[141] Facebook Law Enforcement Guidelines (2010), https://www.eff.org/files/filenode/social_network/facebook2010_sn_leg-doj.pdf
[142] *See* Appendix A.

in constitutionally protected protest or activism. Moreover, it is insufficient for companies to merely discourage "surveillance" without expanding on what that definition means, as Facebook and Instagram do. Twitter's definition of "surveillance" is both expansive and comprehensive, and should be a model for other companies attempting to enact similar prohibitions.

Second, currently only Facebook prevents undercover or "fake" accounts as a by-product of its real-name policy. This policy has been controversial and roundly criticized in other spheres—for instance, the LGBT community has pointed out that in the transgender community the "real-name" policy places an additional burden by preventing individuals from using adopted names or pseudonyms that may well be a part of their identity—such as drag queens' stage names.[143] It therefore doesn't seem like a practicable move to encourage other platforms to adopt a similar policy. However, platforms can require that police departments who create undercover or "fake" accounts demonstrate proof of internal policies or oversight governing the use of such accounts. As with Austin's police department, requiring that police departments only use undercover accounts when there is a clear suspicion or nexus to criminal activity may cut down on the use of such accounts for pure surveillance. This may not necessarily curb the use of this technique, but it will provide some accountability for police by minimally mandating that such surveillance conform to documented standards. Platforms could also enact a blanket ban on communicating with users known to be minors through undercover or fake accounts.

It is promising that all four platforms' general policy is to notify users of requests for their data. Platforms should consider adopting Twitter and Snap's additional safeguards: namely, explicitly providing users with a copy of the request, and allowing users seven days to contest the request in

[143] Miz Cracker, *Facebook's "Real Name" Fix Isn't a Fix At All,* Slate (Nov. 4, 2015), https://slate.com/human-interest/2015/11/facebook-real-name-policy-the-changes-for-trans-people-drag-queens-and-others-arent-as-meaningful-as-they-seem.html.

court before complying. All four platforms are also somewhat vague about what happens after a non-disclosure order expires: namely, whether users are eventually notified that their data was requested even if platforms are initially barred by court order from letting them know. A final safeguard would be for platforms to explicitly state that they will notify affected users following the expiration of a non-disclosure order.

Currently, Snap appears to provide the most specifics in terms of how long different types of data is stored and when such data is permanently deleted, but this is a low bar considering the level of vagueness inherent in all four platform policies. Platforms should provide a clear accounting of retention policies for each category of information they collect, rather than merely specifying that such retention times vary. Platforms should also state with specificity the precise types of information collected by them that are available to law enforcement, such as information on user interactions with other accounts. Finally, while Twitter and Snapchat make clear that a user's deletion of individual content commonly means it is also permanently deleted from the site, Facebook is vague about whether such deletion permanently removes such data and whether such deleted information can (and will) be released to law enforcement. Facebook (and by extension, Instagram) would therefore benefit from more clarity in their policies on exactly what sorts of data users can expect to be permanently deleted, and when.

Finally, though transparency reports currently provide a significant amount of information on the number of legal requests for user information, no platform explicitly addresses other forms of surveillance in its transparency reporting, leaving users to hunt for such information in piecemeal blog posts or press releases. Facebook and Twitter could provide a list of third-party developers found to have been violating their prohibitions on using data for surveillance; while Facebook could provide numbers and locations of fake law enforcement profiles disabled. A more comprehensive accounting of such abuses and the platforms' response thereof would be immensely helpful, both in

terms of providing activists and organizers with specific information on the platforms' role in countering and curbing surveillance; as well as an accountability metric to law enforcement and the entities that provide surveillance services or software to them, indicating that such abuses will not be tolerated and will be publicly disclosed by the platforms themselves.

CONCLUSION

Social media is a key player in every modern movement to change the world, from #MeToo and the Women's Marches, to the Arab Spring, to Black Lives Matter. Its potential to connect people and spread information across borders in an instant has revolutionized how people protest, communicate, and organize. However, the same networks and communication powers that make social media such a potent tool for social justice make it an equally valuable tool for law enforcement seeking to constrain or limit the power of these movements. As this paper demonstrates, not only does social media surveillance chill the legitimate expression of speech and political dissent, its impact is most sharply felt by communities of color who have borne the brunt of over-policing and mass incarceration for generations. These struggles are not new, but technology imbues them with a new urgency for policy and legal solutions, in order to ensure that the constitutional freedoms that underpin our democracy are not threatened by expanding surveillance.

Yet, even as we wait for the law to catch up to evolving notions of digital privacy, social media platforms themselves can play a significant—and often overlooked—role in preserving the balance between security and civil liberties, by ensuring their platforms are not exploited for surveillance in the absences of legitimate suspicions of criminal activity. This paper attempts to demonstrate some ways in which platforms with a legitimate commitment to user privacy and social justice can constrain the use of their data for unlawful surveillance. Ultimately, if companies wish to

stick to their stated missions of "bringing the world together," they cannot ignore the potential of their platforms to be subverted into a mechanism to tear communities apart.

Appendix A: Table of Relevant Privacy, Data Retention, and Law Enforcement Policies

	Twitter	Facebook	Instagram	Snapchat
User information available to non-platform users?	Yes- can pull up user with direct profile link, no login or user account required.	No- direct search without logging in shows limited user information.	Yes- can pull up user with direct profile link, no login or user account required.	No. Users must have a Snapchat account to view others' stories.
Privacy controls available to users?	Users can set their full profile to private, enabling only approved followers to access data; otherwise, profiles are fully public. Allows blocking and muting of other users.	Individuals can change privacy settings for each piece of content posted except for current profile picture and profile name. Users not "friended" see limited profile information unless user specifically sets privacy settings to "public."	Users can set their full profile to private, enabling only approved followers to access data. Allows blocking and muting of other users. No content-specific privacy controls available for general posts, though Instagram has just launched a "close friends" option for Instagram stories, allowing only approved followers to see specific stories.	By default, only "friends" added by user can contact directly, view posted Stories, or see users' locations on SnapMap. Gives users options to make their accounts public or hide stories from specific users, even friends.
Guidelines Publicly Available?	Yes.	Yes. Leaked version also available.	Yes.	Yes.
Ban on surveillance?	Yes. Twitter's contract for third party developers explicitly prohibits the use of Twitter Content and information derived from it to 1) any public sector entity (or any entities providing services to such entities) for surveillance purposes, including: investigating or tracking Twitter's users or their Twitter Content; and, tracking, alerting, or other monitoring of sensitive events (including but not limited to protests, rallies, or community organizing meetings); 2) any public sector entity (or any entities providing services to such entities) whose primary function or mission includes conducting surveillance or gathering intelligence; 3) any entity for the purposes of conducting or providing surveillance, analyses or research that isolates a group	"Protect the information you receive from us against unauthorized access, use, or disclosure. For example, don't use data obtained from us to provide tools that are used for surveillance. "https://developers.facebook.com/policy/	"Protect the information you receive from us against unauthorized access, use, or disclosure. For example, don't use data obtained from us to provide tools that are used for surveillance.	None found.

	of individuals or any single individual for any unlawful or discriminatory purpose or in a manner that would be inconsistent with our users' reasonable expectations of privacy; 4) any entity to target, segment, or profile individuals based on health (including pregnancy), negative financial status or condition, political affiliation or beliefs, racial or ethnic origin, religious or philosophical affiliation or beliefs, sex life or sexual orientation, trade union membership, data relating to any alleged or actual commission of a crime, or any other sensitive categories of personal information prohibited by law; 5) any entity that you reasonably believe will use such data to violate the Universal Declaration of Human Rights (located at http://www.un.org/en/documents/udhr/), including without limitation Articles 12, 18, or 19.			
Legal Process Requirements	"Twitter responds to valid legal process issued in compliance with applicable law." Disclosure of any non-public information about Twitter users requires a subpoena, court order, or other valid legal process. Contents of communications (such as tweets, direct messages, or photos) require a valid search warrant or equivalent from an agency with proper jurisdiction over Twitter.	A valid subpoena issued in connection with an official criminal investigation is required to compel the disclosure of basic subscriber records (defined in 18 U.S.C. Section 2703(c)(2)), which may include: name, length of service, credit card information, email address(es), and a recent login/logout IP address(es), if available. A court order issued under 18 U.S.C. Section 2703(d) is required to compel the disclosure of certain records or other information pertaining to the account, not including contents of communications, which may include message headers and IP addresses, in	Instagram "disclose[s] account records solely in accordance with ToS and applicable law, including the federal Stored Communications Act. They require a "valid subpoena" to disclose basic subscriber records, which may include: name, length of service, credit card information, email address(es), and any recent login/logout IP address(es), if available; a court order is required to disclose contents of communications, including message headers and IP addresses; and a search warrant is required to access the stored contents of any account, which may include messages, photos, comments, and location information. Please note that a government-issued email address is required to	Complies with SCA; warrant required for location data, message or media content; warrant or court order under 18 U.S.C. §2703(d) required for logs; and warrant, subpoena, or court order under 18 U.S.C. §2703(d) required for basic subscriber information. (Guide at 8-10)

		addition to the basic subscriber records identified above. A search warrant issued under the procedures described in the Federal Rules of Criminal Procedure or equivalent state warrant procedures upon a showing of probable cause is required to compel the disclosure of the stored contents of any account, which may include messages, photos, videos, timeline posts, and location information. We interpret the national security letter provision as applied to Facebook to require the production of only 2 categories of information: name and length of service.	access the Law Enforcement Online Request System. https://help.instagram.com/494561080557017	
What information is available?	"Most Twitter account information is public, so anyone can see it. A Twitter account profile contains a profile photo, header photo, background image, and status updates, called Tweets." Users can also add their location and a short bio. Twitter also collects, uses, and stores information about users' locations, IP addresses, and device settings. If users enable "precise location," Twitter can collect, store, and use your precise location (longitude and latitude); and this information will be associated with your Tweet and findable via API.	Basic subscriber information: user ID number, email address, information about account creation, most recent logins and registered mobile number; Expanded subscriber content: profile contact information, mini-feed, status update history, shares, notes, wall postings, friend and group listings with IDs, future and past events, and video listings. User photos: both uploaded by user and uploaded by others where user is tagged. Group information: BSI about group creator and current status of group, private messages "if retained." IP logs are limited and frequently incomplete. https://www.eff.org/files/filenode/social_network/facebook2010_sn_leg-doj.pdf	Information collected includes "information in or about the content your provide (like metadata) such as the location of a photo or the date a file was created," information about the people, Pages, accounts, hashtags and groups you are connected to and how you interact with them across our Products, such as people you communicate with the most or groups you are part of, information about usage, including "the types of content you view or engage with; the features you use; the actions you take; the people or accounts you interact with; and the time, frequency and duration of your activities. For example, we log when you're using and have last used our Products, and what posts, videos and other content you view on our Products." Unclear what law enforcement can access.	Basic subscriber information (username, email, phone number, account creation date and IP address, and timestamp and IP address of account logins and logouts); logs containing metadata of a user's Snaps, Stories, and Chats, but not the actual content; location data. (Guide at 8-10)
User Notification	Twitter's policy is to notify users of requests for their	Our policy is to notify people who use our service	Our policy is to notify people who use our service of requests	Generally maintains policy

	Twitter or Periscope account information, which includes a copy of the request, as soon as we are able (e.g., prior to or after disclosure of account information) unless we are prohibited from doing so (e.g., an order under 18 U.S.C. § 2705(b)). We ask that any non-disclosure provisions include a specified duration (e.g., 90 days) during which Twitter is prohibited from notifying the user. Exceptions to user notice may include exigent or counterproductive circumstances, such as emergencies regarding imminent threat to life, child sexual exploitation, or terrorism.	of requests for their information prior to disclosure unless we are prohibited by law from doing so or in exceptional circumstances, such as child exploitation cases, emergencies or when notice would be counterproductive. We will provide delayed notice upon expiration of a specific non-disclosure period in a court order and where we have a good faith belief that exceptional circumstances no longer exist, and we are not otherwise prohibited by law from doing so. Law enforcement officials who believe that notification would jeopardize an investigation should obtain an appropriate court order or other appropriate process establishing that notice is prohibited. If your data request draws attention to an ongoing violation of our terms of use, we will take action to prevent further abuse, including actions that may notify the user that we are aware of their misconduct.	for their information prior to disclosure unless we are prohibited by law from doing so or in exceptional circumstances, such as child exploitation cases, emergencies or when notice would be counterproductive. We will provide delayed notice upon expiration of a specific non-disclosure period in a court order and where we have a good faith belief that exceptional circumstances no longer exist, and we are not otherwise prohibited by law from doing so. Law enforcement officials who believe that notification would jeopardize an investigation should obtain an appropriate court order or other appropriate process establishing that notice is prohibited. If your data request draws attention to an ongoing violation of our terms of use, we will take action to prevent further abuse, including actions that may notify the user that we are aware of their misconduct.	of notifying affected Snapchat users "when we receive legal process seeking their records, information, and content." Provides affected users seven days to challenge legal process in court before responding. They do not provide user notice either when prohibited by a court order issued under 18 U.S.C. §2705(b) or similar legal authority; or in the event of an "exceptional circumstance" such as "cases involving child exploitation or the threat of imminent death or bodily injury" (Guide at 3). Policy has been in place since Nov. 2015. https://www.snap.com/en-US/privacy/transparency/
Transparency Reports	Has published since 2012; says it pushes back on requests that are improper; most recently narrowed or didn't disclose information to 46% of government information requests. Does not break down between requests from law enforcement and government agencies generally. Shows emergency records requests; responses for both for US figures is 73-76%. Removal requests: most requests come from Turkey and Romania; received 99 takedown requests from the US, complied with 0.	Yes. Discloses number of requests and account identifiers made pursuant to US legal process and FISA/NSL; discloses number of urgent/expedited requests, provides comprehensive overview of content takedowns and legal requests received internationally.	Does not publish separate report. Facebook doesn't aggregate differences between Facebook and Instagram.	Yes. Discloses number of requests and account identifiers made pursuant to US legal process and FISA/NSL (but no emergency requests); number of requests, account identifiers and emergency requests made pursuant to int'l legal process. Also discloses

	https://transparency.twitter.com/en/information-requests.html			governmental content takedown requests (i.e. content that would otherwise violate ToS) and DMCA copyright takedown notices. High rate of compliance between 86% and 100% for US legal process. https://www.snap.com/en-US/privacy/transparency/
How long is information generally retained?	Twitter retains different types of information for different time periods, and in accordance with our Terms of Service and Privacy Policy. Given Twitter's real-time nature, some information (e.g., IP logs) may only be stored for a very brief period of time. Twitter emphasizes that most data on its platform is public	We store data until it is no longer necessary to provide our services and Facebook Products, or until your account is deleted - whichever comes first. This is a case-by-case determination that depends on things like the nature of the data, why it is collected and processed, and relevant legal or operational retention needs. For example, when you search for something on Facebook, you can access and delete that query from within your search history at any time, but the log of that search is deleted after 6 months. If you submit a copy of your government-issued ID for account verification purposes, we delete that copy 30 days after submission.	Not clear. We retain different types of information for different time periods. Given the volume of real-time content on Instagram, some information may only be stored for a short period of time.	Snap's servers automatically delete a Snap after being viewed by intended recipients. Snap's servers are designed to automatically delete an unopened Snap or message sent directly to a recipient after 30 days and an unopened Snap or message in Group Chat after 24 hours. Snaps sent to the crowdsourced "Our Story" (which is publicly viewable "may be saved for longer periods of time." (unspecified) (guide at 4). Users can also save "sent or unsent Snaps, posted Stories, and photos and videos from their phone's photo gallery in Memories," Snap's cloud-storage

				mechanism, in which case the content is saved until deleted by the user. Snap is vague about the length of retention of location data, stating that such data is collected and retention periods vary.
Policy on Fake/Fraudulent Accounts	Users can use Twitter under a pseudonym "if you prefer not to use your name" and users are allowed to create and manage multiple Twitter accounts. Twitter doesn't require real name use, email verification, or identity authentication. https://help.twitter.com/en/rules-and-policies/twitter-law-enforcement-support#5	People on Facebook are required to use the name they go by in everyday life and must not maintain multiple accounts. Operating fake accounts, pretending to be someone else, or otherwise misrepresenting your authentic identity is not allowed, and we will act on violating accounts. https://www.facebook.com/communitystandards/misrepresentation/	Instagram does not require people to use their real names or identities.	n/a
Encryption Options?	No.	No.	No.	Memories has an encryption option (called "My Eyes Only); if enabled, Snap cannot decrypt stored information.
Is content that has been changed or deleted by users available to law enforcement?	Once an account has been deactivated, there is a very brief period in which we may be able to access account information, including Tweets. Deactivated accounts can be restored within 30 days. Content removed by account holders (e.g., Tweets) is generally not available.	When you delete your account, we delete things you have posted, such as your photos and status updates, and you won't be able to recover that information later. Information that others have shared about you isn't part of your account and won't be deleted.	Not clear. We retain different types of information for different time periods. Given the volume of real-time content on Instagram, some information may only be stored for a short period of time. We do not retain data for law enforcement purposes unless we receive a legally valid preservation request. We also retain information from accounts disabled for terms violations for at least a year to prevent repeat abuse or other term violations.	Generally no, because Snap's servers are designed to automatically delete most user content, they "often cannot retrieve user content except in very limited circumstances. Snap states most user information is deleted if individuals delete their accounts, but "cannot promise that all deletion will occur within a specific

				timeframe," and such deletion may be stalled in response to a valid preservation request. https://www.snap.com/en-US/privacy/privacy-policy/
Exigent circumstances?	Evaluates emergency disclosure requests on a "case-by-case basis in compliance with relevant law;" allows disclosure in "exigent emergenc[ies] that involves the danger of death or serious physical injury to a person."	In responding to a matter involving imminent harm to a child or risk of death or serious physical injury to any person and requiring disclosure of information without delay, a law enforcement official may submit a request through the Law Enforcement Online Request System.	Instagram will respond to exigent requests in matters "involving imminent harm to a child or risk of death or serious physical injury to any person."	Yes: must be submitted by a sworn law enforcement official and must come from an official law enforcement email domain.
Third-party access to data for surveillance.	"We prohibit developers using the Public APIs and Gnip data products from allowing law enforcement — or any other entity — to use Twitter data for surveillance purposes. if developers violate our policies, we will take appropriate action, which can include suspension and termination of access to Twitter's Public APIs and data products." https://blog.twitter.com/developer/en_us/topics/community/2016/developer-policies-to-protect-peoples-voices-on-twitter.html	"Protect the information you receive from us against unauthorized access, use, or disclosure. For example, don't use data obtained from us to provide tools that are used for surveillance.	Instagram's platform policy "Protect the information you receive from us against unauthorized access, use, or disclosure. For example, don't use data obtained from us to provide tools that are used for surveillance. https://www.instagram.com/about/legal/terms/api/	Provides third-party access to API through separate Snap Kit platform; but developers do not have access to "user-identifiable information such as demographic information or friends list." https://www.wired.com/story/snap-kit/ completely blocks third-party apps designed to be used in conjunction with Snapchat after a hacking resulted in 100,000 private snapchat photos being leaked. No explicit ban on using API for surveillance. https://www.lifewire.com/snapchat-blocked-third-party-apps-3485997

				Unclear if policy explicitly bans surveillance as use.
Law enforcement use of platform	n/a	Facebook makes clear that it will "always disable accounts that supply false or misleading profile information or attempt to technically or socially circumvent site privacy measures," and makes clear that Facebook's statement of rights and responsibilities applies to law enforcement use as well.	n/a	n/a