



# THE CASE FOR TRANSPARENCY

HOW SOCIAL MEDIA  
PLATFORM DATA ACCESS  
LEADS TO REAL-WORLD  
CHANGE.

**Naomi Shiffman**  
**Brandon Silverman**

May 2025

Institute for Data,  
Democracy & Politics

---

THE GEORGE WASHINGTON UNIVERSITY

## Executive Summary

This report makes the ‘case for transparency’ by presenting a series of influential social media research efforts that are directly tied to accountability and change across multiple domains:

- **Platform Decision-Making:** Insights from researchers and journalists have led to changes in platform policies, such as improved moderation practices and increased scrutiny over misinformation, hate speech, and state-sponsored influence campaigns.
- **Government and Regulatory Action:** Research efforts have directly informed legal actions and policy decisions, such as discrimination lawsuits, platform regulation efforts, and legislative proposals.
- **Public Awareness:** Investigative journalism and open-source research have helped uncover foreign interference, disinformation campaigns, and radicalization pipelines, leading to public discourse and civic engagement.
- **Academic Understanding:** Transparency initiatives provide empirical data for researchers to study, among other topics, content moderation, advertising, wellbeing, polarization, and manipulation tactics, which informs both scholarly work and public policy.

## Introduction

Over the past decade, social media platforms have moved from offering researchers and other public-interest users large volumes of public data via official voluntary transparency programs to making this data increasingly inaccessible. This tightening increases risks for society and platforms alike. Platform transparency has had significant positive impact for public awareness and civic engagement, academic understanding of the impacts of social media platforms, government and regulatory action to ensure platforms are acting for the benefit of society, and for the platforms themselves, who adjust their policies and enforcement as informed by external researchers and journalists flagging harms.

There is a robust body of computational social science and decades of internet research that have impacted a wide array of industries. In the interest of making the case for transparency, this paper highlights a small number of prominent cases demonstrating the positive impacts of transparency across platform decision-making, regulation, public awareness, and academic understanding, especially connected to voluntary transparency programs that have been shut down.

Perhaps obviously, tying one research or journalistic project to a concrete outcome is difficult. When companies and policymakers make decisions, it is often the result of a lot of information and pressure from many sources, but the cases included in this paper have a fairly clear tie to a specific impact. Cases were selected by a combination of soliciting examples from former platform employees, the research community, journalists and civil society, and pulling well-known examples that made major news. The list is not exhaustive—it is meant to provide a range of examples and start a conversation, as well as kickstart a collection of a larger set of examples. This paper does not mean to downplay other research and advocacy that may have been happening in parallel to the cases featured here, and readers who know of parallel work that is publicly documented are encouraged to share it for inclusion in future work.

## Platform Decision-making

Cases where platforms decided to change their policies, their products or their resources/staffing as a result of information gained through transparency initiatives.

### Case Study #1: Twitter Platform Enforcement following the Russian Invasion of Ukraine (2022)

- **Research:** Following the Russian invasion of Ukraine, independent researchers used the Twitter API to identify coordinated inauthentic behavior in support of Russia. For example, researcher Mark Owen Jones identified a network of fake accounts promoting the hashtag #IStandWithPutin,<sup>1</sup> leading to Twitter taking down that network. Other research work identified coordination to amplify Russian government and embassy accounts.<sup>2</sup>
- **Impact:** While Twitter's blog post on the subject<sup>3</sup> uses oblique language, former Head of Trust & Safety Yoel Roth directly credits independent research in helping the platform to identify and action accounts that were violating its policies.<sup>4</sup>
- **Data sources used:**
  - Official platform programs: Twitter API (now defunct)

### Case Study #2: Research partnerships for state-backed influence operation enforcement (2021)

- **Research:** Twitter regularly worked with independent research organizations leveraging the Twitter API to remove networks of state-backed accounts engaging in influence operations.<sup>5</sup>

---

<sup>1</sup> Collins, Ben, and Natasha Korecki. 2022. "Twitter Bans over 100 Accounts That Pushed #IStandWithPutin." NBC News. March 4, 2022. <https://www.nbcnews.com/tech/internet/twitter-bans-100-accounts-pushed-istandwithputin-rcna18655>.

<sup>2</sup> Purtill, James. 2022. "Twitter Bot Network Amplifying Russian Disinformation about Ukraine War, Researcher Says." ABC News, March 29, 2022. <https://www.abc.net.au/news/science/2022-03-30/ukraine-war-twitter-bot-network-amplifies-russian-disinformation/100944970>.

<sup>3</sup> McSweeney, Sinéad. 2022. "Our Ongoing Approach to the War in Ukraine." Archive.org. 2022. [https://web.archive.org/web/20231211225519mp\\_/https://blog.twitter.com/en\\_us/topics/company/2022/our-ongoing-approach-to-the-war-in-ukraine](https://web.archive.org/web/20231211225519mp_/https://blog.twitter.com/en_us/topics/company/2022/our-ongoing-approach-to-the-war-in-ukraine).

<sup>4</sup> Culliford, Elizabeth. 2022. "Twitter Won't Recommend Russia Govt Accounts, Will Prohibit Some POW Content." Reuters, April 5, 2022. <https://www.reuters.com/technology/twitter-wont-recommend-russia-govt-accounts-will-prohibit-some-pow-content-2022-04-05/>.

<sup>5</sup> "Moderation Research - X Transparency Center." 2018. Archive.org. 2018. <https://web.archive.org/web/20240714160144/https://transparency.x.com/en/reports/moderation-research>.

- **Impact:** For example, in December 2021, its partnership with the Stanford Internet Observatory, the Australian Strategic Policy Institute, and Cazadores de Fake News resulted in the removal of networks originating in Mexico, China, Russia, Tanzania, Uganda, and Venezuela.<sup>6</sup>
- **Data sources used:**
  - Official platform programs: Twitter API (now defunct)

### Case Study #3: Rabbit holes and radicalization on YouTube (2019)

- **Research:** New York Times reporter Kevin Roose wrote an in-depth piece on one man's experience being radicalized via YouTube.<sup>7</sup> It covered the way in which YouTube's product and policy changes starting in 2012 empowered right-wing creators and allowed them to tap into an emerging wave of right-wing populism through the recommendation algorithm. The piece was heavy on anecdotal reporting, but also leveraged YouTube data to trace the path of the subject's video recommendations.
- **Impact:** After contacting YouTube for comment on the piece, a relevant policy update was released a week before the piece was published.<sup>8</sup> The update stated that YouTube would "begin reducing recommendations of borderline content and content that could misinform users in harmful ways."
- **Data sources used:**
  - Manual pulls/scraping: YouTube

### Case study #4: Coordinated behavior across 18 pro-Saif Gaddafi Facebook Pages (2020)

- **Research:** Research conducted by the Stanford Internet Observatory in April 2020 supported Facebook removing a cluster of 18 Pages for inauthentic behavior;<sup>9</sup> these were not taken down for coordinated inauthentic behavior (such as that indicative of state involvement), but they did violate terms of service surrounding amplification through the use of inauthentic accounts (accounts that falsely represent themselves to manipulate their audiences). The Pages were opportunistic in their narratives; while they spent time on nostalgia posts about Muammar Gaddafi, more recently they created posts about COVID-19 designed to denigrate Libya's internationally-recognized Government of National Accord (GNA). The coronavirus posts included claims that coronavirus was a bioweapon created by the CIA. The Pages frequently

---

<sup>6</sup> "Disclosing State-Linked Information Operations We've Removed." 2021. Archive.org. 2021. [https://web.archive.org/web/20240319005736/https://blog.twitter.com/en\\_us/topics/company/2021/disclosing-state-linked-information-operations-we-ve-removed](https://web.archive.org/web/20240319005736/https://blog.twitter.com/en_us/topics/company/2021/disclosing-state-linked-information-operations-we-ve-removed).

<sup>7</sup> Roose, Kevin. 2019. "The Making of a YouTube Radical." The New York Times, June 8, 2019. <https://www.nytimes.com/interactive/2019/06/08/technology/youtube-radical.html>.

<sup>8</sup> The YouTube Team. 2019. "Continuing Our Work to Improve Recommendations on YouTube." Blog.youtube. January 25, 2019. <https://blog.youtube/news-and-events/continuing-our-work-to-improve/>.

<sup>9</sup> Grossman, Shelby, Khadija H, and Renee DiResta. 2020. "Coordinated Behavior across 18 Pro-Saif Gaddafi Facebook Pages." Stanford.edu. April 27, 2020. <https://cyber.fsi.stanford.edu/content/oya-cluster-takedown>.

linked to state media from other countries, including Russia's RT and Sputnik, and referenced reporting from the Russian Foundation for National Values Protection. These Pages frequently shared the exact same messages at the exact same time, and worked to create the impression that Saif Gaddafi had broad popular support among Libyans.

- **Impact:** The network was removed from Facebook.
- **Data sources used:**
  - Official platform program: CrowdTangle (now defunct), Meta Coordinated Inauthentic Behavior archive (hosted on CrowdTangle).

### Case study #5: Genocide in Myanmar (2018)

- **Research:** In 2018, researcher Ray Serrato at Democracy Reporting International examined about 15,000 Facebook posts between June 2016 and August 2017 from supporters of the hardline nationalist Ma Ba The group in Myanmar using the then-active Facebook API. His analysis showed that activity within the anti-Rohingya group, which has 55,000 members, exploded with posts registering a 200% increase in interactions.<sup>10</sup> This analysis and subsequent reporting drew international attention and mounted pressure on Facebook to take action.<sup>11,12</sup>
- **Impact:** Facebook eventually took responsibility for its failure to address hate speech on the platform in Myanmar,<sup>13</sup> and increased its investment in the country and in human rights overall.
- **Data sources used:**
  - Official platform programs: Facebook API (preceded CrowdTangle, was deprecated in 2019)

---

<sup>10</sup> Safi, Libby Hogan Michael. 2018. "Revealed: Facebook Hate Speech Exploded in Myanmar during Rohingya Crisis." The Guardian, April 3, 2018, sec. World news. <https://www.theguardian.com/world/2018/apr/03/revealed-facebook-hate-speech-exploded-in-myanmar-during-rohingya-crisis>.

<sup>11</sup> Gilbert, David. 2018. "Hate Speech Is Still Going Viral on Facebook in Myanmar, despite Zuckerberg's Promises." VICE. April 10, 2018. <https://www.vice.com/en/article/zuckerberg-says-facebook-is-taking-its-myanmar-problem-seriously-activists-say-thats-bs/>.

<sup>12</sup> Stecklow, Steve. 2018. "Why Facebook Is Losing the War on Hate Speech in Myanmar." Reuters. August 15, 2018. <https://www.reuters.com/investigates/special-report/myanmar-facebook-hate/>.

<sup>13</sup> Warofka, Alex. 2018. "An Independent Assessment of the Human Rights Impact of Facebook in Myanmar - about Facebook." About Facebook. November 6, 2018. <https://about.fb.com/news/2018/11/myanmar-hria/>.

## Regulation

Cases where transparency (research, or access to information) informs legislative proposals, regulatory actions (fines, etc.) or compliance monitoring.

### Case Study #1: Airbnb and short-term rental regulation (2015)

- **Research:** A lack of data about short-term housing rentals via Airbnb meant that regulators couldn't assess how these rentals were impacting the housing market. InsideAirbnb,<sup>14</sup> a project created and run by activists and researchers, was created in 2015 by scraping Airbnb listings to create an aggregate data source spanning hundreds of cities. Airbnb argued that hosts primarily were renting out rooms in their homes, or a single unit, and therefore they weren't having a meaningful impact on the housing market. However, the data showed that in fact, there were many hosts that rented out multiple units, up to the many hundreds of units per host in some cities.
- **Impact:** This research led to short-term rental regulation in many cities in order to counter the negative impact on the housing market revealed by the data. In New York City, for example, regulation states that hosts must stay in the same unit as guests, and no more than 2 guests are allowed.<sup>15</sup> InsideAirbnb continues to provide data to New York City and New York State to support continued monitoring and enforcement. In addition, New York State's short-term rental law mandates data-sharing from cities with a short-term rental registry to the state.<sup>16</sup> European cities such as Paris and Amsterdam have also benefited from InsideAirbnb, and the EU recently passed a data-sharing law for short-term rentals as well.<sup>17</sup>
- **Data used:**
  - Manual pulls/scraping: Airbnb data

### Case Study #2: Anti-deepfake pornography regulation (2024)

- **Research:** In April 2024, **ABC News** in Australia used CrowdTangle and YouTube to uncover a scheme to make money by creating faces with AI and then transplanting them onto videos of

---

<sup>14</sup> "Home." n.d. Insideairbnb.com. <https://insideairbnb.com/>.

<sup>15</sup> "Short-Term Rental Registration · NYC311." 2023. Nyc.gov. 2023. <https://portal.311.nyc.gov/article/?kanumber=KA-03559>.

<sup>16</sup> "NY State Senate Bill 2023-S885C." 2023. NYSenate.gov. 2023.

<https://www.nysenate.gov/legislation/bills/2023/S885/amendment/C>.

<sup>17</sup> "Data Collection and Sharing Relating to Short-Term Accommodation Rental Services | Think Tank | European Parliament." 2024. Www.europarl.europa.eu. May 29, 2024.

[https://www.europarl.europa.eu/thinktank/en/document/EPRS\\_BRI\(2023\)739334](https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2023)739334).

women's bodies sourced from social media accounts.<sup>18</sup> The creators then use the fake personas to encourage their followers to join subscription services like Fanvue, an OnlyFans-like platform where they can pay for pornographic images and videos of the AI model.

- **Impact:** This investigation and others helped lead to the passage of an Australian law targeting non-consensual deepfake porn.<sup>19</sup>
- **Data sources used:**
  - Official platform programs: CrowdTangle (now defunct)
  - Manual pulls/scraping: YouTube, PayPal, Instagram and Twitter (now X)

### Case Study #3: Facebook settlement with US DOJ over ad targeting (2016-2019)

- **Research:** In 2016, ProPublica used Facebook's advertising tools to demonstrate that they could be used to advertise housing to specific racial demographics -- a violation of the Fair Housing Act of 1968.<sup>20</sup>
- **Impact:** As a result, in 2019, the Department of Housing and Urban Development sued Facebook for engaging in housing discrimination and violating the Fair Housing Act.<sup>21</sup> Facebook settled with the Department of Justice, including agreeing to change its advertising systems, in June 2022.<sup>22</sup>
- **Data sources used:**
  - Official platform programs: Facebook advertising tools

---

<sup>18</sup> Nguyen, Kevin, and Michael Workman. 2024. "'AI Pimps' and Their Fake Influencers Are Mass-Harvesting Women's Videos to Peddle Porn." ABC News, April 13, 2024. <https://www.abc.net.au/news/2024-04-14/how-ai-pimps-influencers-are-using-womens-videos/103658046>.

<sup>19</sup> Olbrycht-Palmer, Joseph. 2024. "Tough New Penalties for Deepfake Deviants." The Australian. The Australian. August 21, 2024. <https://www.theaustralian.com.au/news/latest-news/senate-passes-law-targeting-nonconsensual-deepfake-porn/news-story/3b8b91a2d04e2ee3a586611d93e3eab4>.

<sup>20</sup> Angwin, Julia, and Terry Parris Jr. 2016. "Facebook Lets Advertisers Exclude Users by Race." ProPublica. ProPublica. October 28, 2016. <https://www.propublica.org/article/facebook-lets-advertisers-exclude-users-by-race>.

<sup>21</sup> "HUD Archives: HUD CHARGES FACEBOOK with HOUSING DISCRIMINATION over COMPANY'S TARGETED ADVERTISING PRACTICES." 2019. Archives.hud.gov. March 28, 2019. <https://archives.hud.gov/news/2019/pr19-035.cfm>.

<sup>22</sup> Isaac, Mike. 2022. "Meta Agrees to Alter Ad Technology in Settlement with U.S." The New York Times, June 21, 2022, sec. Technology. <https://www.nytimes.com/2022/06/21/technology/meta-ad-targeting-settlement.html>.

## Public Awareness

Cases where transparency (research, or access to information) impacts the general public's understanding of an issue, or transparency allows for action in the civic sphere (e.g. elections or accountability for a political candidate)

### Case Study #1: Network of overseas pages interfering with the UK general election (2024)

- **Research:** In June 2024, **ABC News** in Australia uncovered a network of five Facebook pages based in Nigeria that were spreading Russian talking points, as well as content in support of Nigel Farage's Reform UK party, in advance of the 2024 UK elections.<sup>23</sup> Ad library data showed it targeted the ads to British men who were older than 65 years old and were paid for using Nigerian currency.
- **Impact:** The discovery became front page news,<sup>24</sup> with the Conservative's Deputy PM expressing grave concerns about potential social media foreign interference.<sup>25</sup> Meta shut down the pages following ABC News' investigation on July 1, 2024.<sup>26</sup>
- **Data sources used:**
  - Official platform programs: Facebook Ad Library, CrowdTangle (now defunct)

### Case Study #2: Politifact: Debunking the Wayfair child sex-trafficking conspiracy theory (2020)

- **Research:** In July 2020, Politifact reporter Daniel Funke was getting alerted to an inordinate amount of "wayfair" keywords connected to his CrowdTangle lists of notorious misinformers. Using CrowdTangle, Funke searched for "wayfair" with a combination of Boolean search terms in order to hone down his reporting of a child trafficking conspiracy theory, which was then debunked.<sup>27</sup>

---

<sup>23</sup> Workman, Michael, and Nguyen, Kevin. 2024a. "UK Conservatives Say ABC Analysis That Points to Foreign Interference Operation 'Highly Alarming.'" ABC News, June 29, 2024. <https://www.abc.net.au/news/2024-06-29/uk-election-pro-russian-facebook-pages-coordinating/104038246>.

<sup>24</sup> "Sunday Times Front Page 30th of June 2024 - Tomorrow's Papers Today!" 2024. Tomorrow's Papers Today! June 29, 2024. <https://www.tomorrowpapers.co.uk/sunday-times-front-page-2024-06-30/>.

<sup>25</sup> Workman, Michael, and Nguyen, Kevin. 2024a.

<sup>26</sup> Workman, Michael, and Nguyen, Kevin. 2024b. "Facebook Pages Linked to Foreign Interference in UK Elections Taken down after ABC Investigation." [abc.net.au. ABC News. July 2024. https://www.abc.net.au/news/2024-07-01/meta-shuts-down-pro-russian-facebook-pages-in-uk-elections/104045286](https://www.abc.net.au/news/2024-07-01/meta-shuts-down-pro-russian-facebook-pages-in-uk-elections/104045286).

<sup>27</sup> Funke, Daniel. 2020. "How the Wayfair Child Sex-Trafficking Conspiracy Theory Went Viral." Politifact. July 15, 2020. <https://www.politifact.com/article/2020/jul/15/how-wayfair-child-sex-trafficking-conspiracy-theor/>.

- **Impact:** The Wayfair conspiracy was publicly debunked, helping address the spread of the conspiracy theory on social media platforms.
- **Data sources used:**
  - Official platform programs: CrowdTangle (now defunct)

### Case study #3: Russian Invasion of Ukraine (2022)

- **Research:** Following the 2022 Russian invasion of Ukraine, on Facebook, the DFRLab exposed a massive network of fake pages, groups, and user profiles set up to undermine the West's confidence in Ukraine.<sup>28</sup> Meta's review of these assets, along with others separately identified by German media, led to the discovery of a much larger network consisting of 1,633 accounts, 703 pages, twenty-nine Instagram profiles, and one Facebook group. The collaboration between the DFRLab, media organizations, and the platform effectively countered Russian attempts to manipulate the information landscape and rationalize war crimes and human rights violations in Ukraine. The DFRLab successfully verified and geolocated more than 130 videos of these attacks on Kharkiv, including the targeting of schools, in the report Two Weeks That Shook Kharkiv.<sup>29</sup>
- **Impact:** This report eventually led to Stanford's Starling Lab filing a submission of evidence of war crimes to the International Criminal Court,<sup>30</sup> developing and submitting a cutting-edge cryptographic dossier to preserve multimedia evidence collected by the DFRLab.<sup>31</sup>
- **Data sources used:**
  - Official platform programs: CrowdTangle (now defunct)

### Case study #4: EU Sanctions against Russia Today (2022)

- **Research:** The Lumen database is an independent research project studying takedown notices along with other legal removal requests and demands concerning online content.<sup>32</sup> It collects and analyzes requests to remove material from the web, with the goal of facilitating research about the different kinds of complaints and requests for removal--both legitimate and questionable--that are being sent to Internet publishers, search engines, and service providers, and to provide as much transparency as possible about the "ecology" of such notices, in terms of who is sending them and why, and to what effect. In March 2022, legal researchers identified

---

<sup>28</sup> DFRLab. 2022. "Russia-Based Facebook Operation Targeted Europe with Anti-Ukraine Messaging." Medium. DFRLab. September 27, 2022. <https://medium.com/dfrlab/russia-based-facebook-operation-targeted-europe-with-anti-ukraine-messaging-389e32324d4b>.

<sup>29</sup> Woolbright, Jocelyn. 2022. "Partnering with Civil Society to Track Internet Shutdowns with Radar Alerts and API." The Cloudflare Blog. December 15, 2022. <https://blog.cloudflare.com/partnering-with-civil-society-to-track-shutdowns/>.

<sup>30</sup> Sheldon, Michael J, and Andy Carvin. 2022. "Two Weeks That Shook Kharkiv: A Geographic Tour of Russia's Attacks on Civilian Neighborhoods in Ukraine's Second-Largest City." Atlantic Council. April 4, 2022. <https://www.atlanticcouncil.org/blogs/new-atlanticist/two-weeks-that-shook-kharkiv/>.

<sup>31</sup> Horowitz, Julia. 2022. "Exclusive: A Crypto-Based Dossier Could Help Prove Russia Committed War Crimes." CNN. June 10, 2022. <https://www.cnn.com/2022/06/10/tech/ukraine-war-crimes-blockchain/index.html>.

<sup>32</sup> "About :: Lumen." 2021. Lumendatabase.org. 2021. <https://lumendatabase.org/pages/about>.

an EU takedown request for Russia Today and Sputnik -- two Russian state media sources.<sup>33</sup> Per legal analysts, the request required 1) search engines to delist all their content, and 2) social media firms to delete posts by individuals which reproduce any of their content, and represented a radical interpretation of sanctions regulation.<sup>34</sup> The EU did not publicly announce this major change, and the public was only made aware of it through the Lumen database.

- **Impact:** The radical interpretation of sanctions regulation revealed by the Lumen database documentation both highlighted an important legal shift that was important for the public and legal analysts to understand.
- **Data sources used:**
  - Third party databases: Lumen database

### Case study #5: Cloudflare internet shutdown monitoring (2021)

- **Research:** In 2021, Cloudflare (a major provider of website security services, among other services) launched Radar Internet shutdown alerts to provide alerts on when it has detected significant drops in traffic with the hope that the information is used to document, track, and hold institutions accountable for these human rights violations.<sup>35</sup>
- **Impact:** Many human rights organizations have leveraged the tool to fight against internet shutdowns.<sup>36</sup> For example, Access Now's #KeepItOn coalition, which represents more than 280 organizations from 105 countries across the globe, uses the tool to both find out about and verify internet shutdowns -- it increases the speed with which they can mobilize against shutdowns worldwide.<sup>37</sup>
- **Data sources used:**
  - Official platform programs: Cloudflare Radar

---

<sup>33</sup> "Government Request Removal Complaint to Google :: Notices :: Lumen." 2022. Lumendatabase.org. March 4, 2022. <https://lumendatabase.org/notices/26927483#>.

<sup>34</sup> Team, Lumen Database. 2022. "Notice in Lumen Reveals That EU Mandated Removal of Russian State-Sponsored News from Google's Search Engine." Medium. March 10, 2022. <https://lumendatabase-org.medium.com/notice-in-lumen-reveals-that-eu-mandated-removal-of-russian-state-sponsored-news-from-googles-ad8936a46607>.

<sup>35</sup> Woolbright, Jocelyn. 2022. "Partnering with Civil Society to Track Internet Shutdowns with Radar Alerts and API." The Cloudflare Blog. December 15, 2022. <https://blog.cloudflare.com/partnering-with-civil-society-to-track-shutdowns/>.

<sup>36</sup> Sousa Botto, João. 2021. "Working with Those Who Protect Human Rights around the World." The Cloudflare Blog. July 29, 2021. <https://blog.cloudflare.com/working-with-those-who-protect-human-rights-around-the-world/>.

<sup>37</sup> "KeepItOn: Fighting Internet Shutdowns around the World." n.d. Access Now. <https://www.accessnow.org/campaign/keepiton/>.

## Academic Understanding

Cases where transparency (research, or access to information) impacts our empirical understanding of an issue.

### Case Study #1: Ad Transparency and the Ad Observatory (2022)

- **Research:** Researcher Laura Edelson at NYU wanted to answer the questions, “What are the systems underlying social media platforms doing? Are they doing what the people who built them say they’re doing?”. In order to answer this question with relation to ads, her team built the Ad Observatory<sup>38</sup>, a database to which platform users could donate data about which political ads they were seeing on their Facebook and Instagram feeds using Ad Observer<sup>39</sup>, a proprietary tool. Researchers also used the Facebook Ad Library<sup>40</sup> to compare with Ad Observatory results in order to compare which ads were labeled as political in the Ad Library vs. which were missed, but captured in the Ad Observatory.
- **Impact:** The resulting data supported research on political ad spending on Facebook and Instagram<sup>41</sup>, on the effectiveness of Facebook’s political ad policy enforcement,<sup>42</sup> a proposal for a digital ad transparency standard,<sup>43</sup> research on foreign actors using Facebook ads to spread election disinformation,<sup>44</sup> and more. This last piece of research in particular was heavily covered in the media.<sup>45</sup> The next stage of their research will be to create more empirical bases for measuring what the systems underlying social media platforms are doing.
- **Data sources used:**
  - Donated data: Ad Observatory
  - Official platform programs: Facebook Ad Library

---

<sup>38</sup> “Ad Observatory by NYU Tandon School of Engineering.” n.d. Adobservatory.org. <https://adobservatory.org/>.

<sup>39</sup> “Ad Observer.” 2025. Ad Observer. 2025. <https://adobserver.org/>.

<sup>40</sup> “Ad Library.” n.d. [www.facebook.com/ads/library](https://www.facebook.com/ads/library).

<sup>41</sup> “Abortion and Trump Top Topics for Federal Candidate Political Ad Spending on Facebook and Instagram.” 2022. [Cybersecurityfordemocracy.org](https://cybersecurityfordemocracy.org). 2022. <https://cybersecurityfordemocracy.org/abortion-and-trump-top-topics-for-federal-candidate>.

<sup>42</sup> Pochat, Victor Le, Laura Edelson, Tom Van Goethem, Wouter Joosen, Damon McCoy, and Tobias Lauinger. 2022. “An Audit of Facebook’s Political Ad Policy Enforcement.” [www.usenix.org](https://www.usenix.org/conference/usenixsecurity22/presentation/lepochat). 2022. <https://www.usenix.org/conference/usenixsecurity22/presentation/lepochat>.

<sup>43</sup> Edelson, Laura, Jason Chuang, Erika Franklin Fowler, Michael M. Franz, and Travis Ridout. 2021. “A Standard for Universal Digital Ad Transparency.” [Knightcolumbia.org](https://knightcolumbia.org/content/a-standard-for-universal-digital-ad-transparency). December 9, 2021. <https://knightcolumbia.org/content/a-standard-for-universal-digital-ad-transparency>.

<sup>44</sup> Edelson, Laura, Tobias Lauinger, and Damon McCoy. 2025. “CSDL | IEEE Computer Society.” [Computer.org](https://www.computer.org/csdl/proceedings-article/sp/2020/349700b096/1j2Lg6YVmko). 2025. <https://www.computer.org/csdl/proceedings-article/sp/2020/349700b096/1j2Lg6YVmko>.

<sup>45</sup> Scott, Mark. 2020. “Stealth Political Ads Flourish on Facebook.” [POLITICO](https://www.politico.com/news/2020/03/06/stealth-political-ads-flourish-on-facebook-122539). Politico. March 6, 2020. <https://www.politico.com/news/2020/03/06/stealth-political-ads-flourish-on-facebook-122539>.

## Case study #2: Hate speech and social media in Sri Lanka (2018)

- **Research:** In 2018, a wave of mob violence against Muslims in Sri Lanka was gaining steam, in part due to inflammatory posts on Facebook and WhatsApp.<sup>46</sup> Researcher Sanjana Hattotuwa published policy briefs<sup>47</sup> and academic papers<sup>48</sup> analyzing the way hate speech spread on Facebook and Twitter specifically in Sri Lanka in relation to this wave of violence and the subsequent constitutional crisis, including recommendations for countering hate speech both for platforms and civil society.
- **Impact:** This research increased academic understanding of the spread of hate speech on social media platforms.
- **Data sources used:**
  - Official platform programs: Twitter API (now defunct), CrowdTangle (now defunct)

## Case study #3: Leveraging social media data to predict of the onset of depression (2013)

- **Research:** Researcher Munmun De Choudhury was investigating the impact of social media on mental health. She used a crowdsourced set of Twitter users who reported a diagnosis of clinical depression, and pulled their account data via the Twitter API, to inform her research. She found that social media could be used to effectively predict the onset of depression in Twitter users.<sup>49</sup> Since publishing this paper (the first in its field), she has continued to publish prolifically on the topic of social media and mental health.

---

<sup>46</sup> Goel, Vindu, Hari Kumar, and Sheera Frenkel. 2018. "In Sri Lanka, Facebook Contends with Shutdown after Mob Violence (Published 2018)." The New York Times, March 8, 2018, sec. Technology.

<https://www.nytimes.com/2018/03/08/technology/sri-lanka-facebook-shutdown.html>.

<sup>47</sup> Hattotuwa, Sanjana. 2018. "Digital Blooms: Social Media and Violence in Sri Lanka." November 2018.

[https://toda.org/assets/files/resources/policy-briefs/t-pb-28\\_sanjana-hattotuwa\\_digital-blooms-social-media-and-violence-in-sri-lanka.pdf](https://toda.org/assets/files/resources/policy-briefs/t-pb-28_sanjana-hattotuwa_digital-blooms-social-media-and-violence-in-sri-lanka.pdf).

<sup>48</sup> Hattotuwa, Sanjana, and Shilpa Samaratunge. 2014. "Liking Violence - a Study of Hate Speech on Facebook in Sri Lanka." Policycommons.net, September. <https://policycommons.net/artifacts/1545511/liking-violence/2235320/>.

<sup>49</sup> De Choudhury, Munmun, Michael Gamon, Scott Counts, and Eric Horvitz. 2021. "Predicting Depression via Social Media." Proceedings of the International AAAI Conference on Web and Social Media 7 (1): 128–37.

<https://doi.org/10.1609/icwsm.v7i1.14432>.

- **Impact:** De Choudhury’s research has been foundational in the field of wellbeing-related computational social science.<sup>50</sup> It has been covered extensively in the media,<sup>51,52,53,54,55,56,57,58</sup> and informed the U.S. Surgeon General’s advisory on social media and youth mental health.<sup>59</sup>
- **Data sources used:**
  - Official platform programs: Twitter API (now defunct)

#### Case study #4: Dynamic Debates: An Analysis of Group Polarization Over Time on Twitter (2010)

- **Research:** Researchers Sarita Yardi and Danah Boyd were interested in the dynamics of group polarization on Twitter, which was relatively new at the time. Using the Twitter API, Yardi and Boyd captured 30,000 tweets about the shooting of George Tiller, a late-term abortion doctor, and the subsequent conversations among pro-life and pro-choice advocates. They found that replies between users with similar viewpoints strengthen group identity, while those between users with a difference of opinion reinforce in-group and out-group affiliation. Their results show that while users are exposed to broader viewpoints than they were before using Twitter, they are limited in their ability to engage in meaningful discussion.<sup>60</sup>
- **Impact:** This was one of the earliest papers about group polarization on social media. It is among the foundational papers in the research of social networks.
- **Data sources used:**
  - Official platform programs: Twitter API (now defunct)

---

<sup>50</sup> “Munmun de Choudhury.” 2020. Google.com. 2020.

[https://scholar.google.com/citations?view\\_op=list\\_works&hl=en&hl=en&user=Z9Pfp6UAAAAJ](https://scholar.google.com/citations?view_op=list_works&hl=en&hl=en&user=Z9Pfp6UAAAAJ).

<sup>51</sup> Schwartz, Casey. 2020. “Is Everybody Doing ... OK? Let’s Ask Social Media.” The New York Times, October 12, 2020, sec. Style. <https://www.nytimes.com/2020/10/12/style/self-care/social-media-.html>.

<sup>52</sup> Beck, Julie. 2016. “What Instagrams Look like in Food Deserts.” The Atlantic. March 1, 2016.

<https://www.theatlantic.com/health/archive/2016/03/the-instagramms-of-food-deserts/471540/>.

<sup>53</sup> “The TikTok Spiral, Part 2: Stuck. Together.” 2025. WSJ. 2025. <https://www.wsj.com/podcasts/google-news-update/the-tiktok-spiral-part-2-stuck-together/63E5F70A-ED88-4B52-A27E-66E00C265FB8>.

<sup>54</sup> “People Fixing the World - Predicting Suicide - BBC Sounds.” 2019. BBC. 2019.

<https://www.bbc.co.uk/sounds/play/w3cswvsg>.

<sup>55</sup> Chen, Angus. 2016. “Instagramming in Black and White? Could Be You’re Depressed.” NPR. August 24, 2016.

<https://www.npr.org/sections/health-shots/2016/08/24/490941032/instagramming-in-black-and-white-could-be-youre-depressed>.

<sup>56</sup> Singer, Natasha. 2015. “Love in the Time of Twitter.” Bits Blog. February 18, 2015.

<https://archive.nytimes.com/bits.blogs.nytimes.com/2015/02/13/love-in-the-times-of-twitter/>.

<sup>57</sup> “Can Social Media Help Direct Mental Health Aid?” 2016. NBC News. March 14, 2016.

<https://www.nbcnews.com/tech/social-media/can-social-media-help-direct-mental-health-aid-n53801>.

<sup>58</sup> Singer, Natasha. 2014. “Risks in Using Social Media to Spot Signs of Mental Distress (Published 2014).” The New York Times, December 26, 2014, sec. Technology. <https://www.nytimes.com/2014/12/27/technology/risks-in-using-social-posts-to-spot-signs-of-distress.html>.

<sup>59</sup> “Social Media and Youth Mental Health -- the U.S. Surgeon General’s Advisory.” 2023. U.S. Department of Health and Human Services. 2023. <https://www.hhs.gov/sites/default/files/sg-youth-mental-health-social-media-advisory.pdf>.

<sup>60</sup> Yardi, Sarita, and Danah Boyd. 2010. “Dynamic Debates: An Analysis of Group Polarization over Time on Twitter.” Bulletin of Science, Technology & Society 30 (5): 316–27. <https://doi.org/10.1177/0270467610380011>.

## Conclusion

While this paper lists only some of the most impactful examples of the ways platform transparency contributes to public awareness, regulation, academic understanding, and platform decision-making, there are thousands more examples across many more platforms—soon to be documented in a living archive. These case studies show that successful collaborations between researchers, journalists, policymakers, advocacy groups and the platforms themselves lead to real-world impact.

## Acknowledgements

Thanks to IDDP for funding in support of this paper. Thanks to Anna Lenhart and Rebekah Tromble for the collaboration and support. Thanks to the following researchers, academics, journalists, activists, and former social media platform staff for providing interviews or case studies: Dylan Baker, Munmun de Choudhury, Murray Cox, Ben Decker, Renée DiResta, Laura Edelson, Yael Eisenstadt, Sanjana Hattotuwa, Rose Jackson, Daphne Keller, Rasto Kuzel, Helena Larrauri, Jacqueline Malaret, Chris Murray, Kevin Nguyen, John Perrino, Kevin Roose, Hilary Ross, Yoel Roth, Ray Serrato, and Michael Workman.

IDDP launched in 2019 with the support of the John S. and James L. Knight Foundation. IDDP's mission is to help the public, journalists and policymakers understand digital media's influence on public dialogue and opinion, and to develop sound solutions to disinformation and other ills that arise in these spaces.

**Institute for Data,  
Democracy & Politics**

---

**THE GEORGE WASHINGTON UNIVERSITY**