Article

Suspecting with Data

Mary D. Fan†

Our pooled consumer big data, such as the pictures we post or the location history and keyword search trails we leave, are generating new ways to solve crimes. Much of the commentary on big data search strategies such as keyword, geofence, and facial recognition searches fixate on Fourth Amendment search and seizure issues rather than evidentiary safeguards. This Article breaks new ground by framing evidentiary guardrails for big data searches to reduce the harms of erroneous arrests, redress secrecy, and counteract the mystique of machine infallibility.

Advancing beyond over-reliance on Fourth Amendment doctrine, this Article illuminates how evidence law and procedures are better suited to address concerns over inaccuracy, overbreadth, and opacity surrounding big data search strategies. The Article offers three proposals. First is requiring corroboration before big data search strategies can be the basis to arrest or convict a person, thus updating the concept of probable cause for changing technologies of proof. Second are pretrial notice, disclosure,

† Jack R. MacDonald Endowed Chair, University of Washington School of Law. mdfan@uw.edu. I am grateful to the participants at the Evidence Workshop at Vanderbilt Law School, especially commentator Erin Murphy and founder Ed Cheng, for generative feedback. I also am grateful to the participants at the 17th Annual Why Law Matters Conference (sponsored by the Federal Defenders for the Central District of Illinois and the University of Illinois), the Innocence Network Conference on "AI, Surveillance, and Wrongful Convictions," and participants at training conferences for the Washington Association of Criminal Defense Lawyers, the Oregon Criminal Defense Lawyers Association, and the Idaho Association of Criminal Defense Lawyers. Thanks also to Ronald J. Allen, Sarah Brown-Schmidt, Kiel Brennan-Marquez, Elodie Currier, Cindy Fester, Patrick Keenan, and Emily Spottswood for valuable insights and suggestions. I am especially grateful to outstanding Articles Editor Aidan McNamara and the team at the *Minnesota Law Review*, including Editor-in-Chief Callan Showers, Managing Editors Mark Hager, Samuel Makikalli, and Tiger Rost, for excellent editing and suggestions.

and reliability hearings to pierce the secrecy surrounding the use of big data search strategies and to permit effective defense challenges. Third is deploying expert witnesses on the reliability concerns surrounding the evidentiary fruits of big data analytical techniques to correct the mystique of machine infallibility and the risk of factfinders overweighing match evidence.

TABLE OF CONTENTS

INTRODUCTION

For Harvey Eugene Murphy, Jr., his nightmare began when a company's artificial intelligence-powered facial recognition program erroneously matched images of him scraped from online databases to an armed robber caught on camera.[1] For Jorge Molina, his wrongful arrest arose when a geofence search revealed that a cell phone logged into his accounts was present at the time of a murder; police rushed to arrest him, ignoring evidence that his abusive stepfather used his phone and car.[2] For the Diol family, their deaths arose from a mistaken identification from a distraught teen using the "Find my iPhone" feature to locate his stolen phone.[3] The cold case murders of the Diols were ultimately solved by another big data strategy: A keyword warrant revealing who googled their address before killers set their home afire.[4] These stories show the perils, power, and proliferation of largely unregulated investigative strategies drawing on big data and digital trails that are so pervasive they are used by private persons and businesses, as well as by the police.

Investigative strategies drawing on big data searches and digital trails take different approaches but share some theoretically and pragmatically important features. First is opaque private collection and control of the data for usually commercial reasons.[5] Second is the sharp deviation from the original use case for the data collected by commercial entities when deployed to identify unknown perpetrators of a crime.[6] Third are the

---

1. Plaintiff's Original Petition at 4–5, Murphy v. Essilorluxottica USA Inc., No. 2024-03265 (125th Dist. Ct., Harris Cnty., Tex. Jan. 18, 2023) (on file with the Minnesota Law Review) [hereinafter *Murphy* Complaint].

2. Complaint at 6–7, Molina v. Avondale, No. CV2019-015311 (Maricopa Cnty., Super. Ct., Ariz. Dec. 12, 2019) (on file with the Minnesota Law Review) [hereinafter *Molina* Complaint]; Jon Schuppe, *Google Tracked His Bike Ride Past a Burglarized Home. That Made Him a Suspect*, NBC NEWS (Mar. 7, 2020), https://www.nbcnews.com/news/us-news/google-tracked-his-bike-ride-past-burglarized-home-made-him-n1151761 [https://perma.cc/F8QY-87BX].

3. *Suspects Kevin Bui, Gavin Seymour Appear in Court for 2020 Denver Fire that Killed 5 People*, CBS NEWS (Nov. 12, 2021), https://www.cbsnews.com/colorado/news/suspects-kevin-bui-gavin-seymour-appear-in-court-for-2020-denver-fire-that-killed-5-people [https://perma.cc/23PN-N6CG].

4. Response to Motion to Suppress Evidence Unlawfully Obtained (Cellphone Data) at 2, People v. Seymour, No. 2021CR20001 (Dist. Ct., Denver Cnty., Colo. Aug. 12, 2022) (on file with the Minnesota Law Review) [hereinafter Response to Motion to Suppress Evidence (*Seymour*)].

5. *See* discussion *infra* Part I.B.1.

6. *See* discussion *infra* Part I.B.2.

unknown or disparate risks of error when the data is coopted to identify potentially unknown suspects, including risks that may vary based on geography, race, ethnicity, complexion, and access to resources.[7] Another overarching challenge for these varied big data search strategies is the predominant focus on whether and how the Fourth Amendment's prohibition against unreasonable searches and seizures apply when debating how to curtail potential harms.[8]

Breaking new ground, this Article advances beyond Fourth Amendment fetishism and stalemates, and theorizes why evidence law protections and procedures are better suited to address the concerns posed by such big data search strategies. Though the Supreme Court has sometimes contorted the fifty-four words of the Fourth Amendment to address investigative strategies posed by advancing technology, many gaps and open questions remain hard to fill by constitutional text.[9] Courts are

---

7. *See* discussion *infra* Part I.B.3.

8. *Compare, e.g.*, Mary D. Fan, *Big Data Searches and the Future of Criminal Procedure*, 102 TEX. L. REV. 877, 883–86, 925–31 (2024) (advancing beyond originalism and updating Fourth Amendment particularity and overbreadth analyses for big data searches), *with* Haley Amster & Brett Diehl, Note, *Against Geofences*, 74 STAN. L. REV. 385, 433–37 (2022) (arguing geofence warrants violate the Fourth Amendment), *and* Chelsa Camille Edano, Comment, *Beware What You Google: Fourth Amendment Constitutionality of Keyword Warrants*, 97 WASH. L. REV. 977, 993–99 (2022) (arguing keyword warrants violate Fourth Amendment), *and* Reed Sawyers, *For Geofences: An Originalist Approach to the Fourth Amendment*, 29 GEO. MASON L. REV. 787, 810–16 (2022) (arguing the Fourth Amendment does not apply to geolocation data), *and* Christopher Slobogin, *Suspectless Searches*, 83 OHIO ST. L.J. 953, 959–62 (2022) (arguing geofence warrants pass Fourth Amendment muster, with limitations), *and* Matthew E. Cavanaugh, Note, *Somebody's Tracking Me: Applying Use Restrictions to Facial Recognition Tracking*, 105 MINN. L. REV. 2443, 2495–98, 2500–04 (2021) (arguing facial recognition tracking of seven days or more constitutes a search and may be analyzed for reasonableness), *and* Andrew Guthrie Ferguson, *Facial Recognition and the Fourth Amendment*, 105 MINN. L. REV. 1105, 1141–61 (2021) (applying Fourth Amendment to facial recognition technology), *and* Note, *Geofence Warrants and the Fourth Amendment*, 134 HARV. L. REV. 2508, 2514–20 (2021) [hereinafter Note, *Geofence Warrants and the Fourth Amendment*] (analyzing Google's three-step framework for geofence warrants).

9. *See, e.g.*, Matthew Tokson, *The Aftermath of* Carpenter*: An Empirical Study of Fourth Amendment Law, 2018–2021*, 135 HARV. L. REV. 1790, 1791 (2022) ("Scholars and lower courts have tried to guess at what the law of Fourth Amendment searches will be going forward—and have reached different, contradictory conclusions."); Paul Ohm, *The Many Revolutions of* Carpenter, 32 HARV. J.L. & TECH. 357, 383, 394–408 (2019) (parsing numerous unanswered technological search questions in the wake of the *Carpenter* decision).

split as to whether the Fourth Amendment even applies to regulate tactics like obtaining geofence location data for the brief timespan of a crime, keyword search terms shared with Google, or scans of facial images scraped from the Internet.[10] Under the third-party exposure doctrine, there is no reasonable expectation of privacy protectable by the Fourth Amendment in information shared with a third party, unless a court is willing to extend the Supreme Court's decision in *Carpenter v. United States* beyond its focus on pervasive data searches of seven days or more of location information.[11]

Puzzlingly overshadowed by the glamour of constitutional criminal procedure, evidence law is theoretically and pragmatically better suited to address the risks of error posed by big data search strategies.[12] Improving accuracy in fact-finding and fairly allocating the risks of error are core concerns of evidence law, procedure, and theory.[13] In contrast, privacy against

10. *Compare, e.g.*, United States v. Chatrie, 107 F.4th 319, 330–32 (4th Cir. 2024) (holding that obtaining two hours of geofence location data around the time and place of a crime is not a Fourth Amendment-regulated search because there is no reasonable expectation of privacy in information shared with third parties), *reh'g en banc granted*, 2024 WL 4648102 (4th Cir. Nov. 1, 2024), *and* State v. Contreras-Sanchez, 5 N.W.3d 151, 163–164 (Minn. Ct. App. 2024) (rejecting analogy between geofence warrants and general warrants and holding that geofence warrants may be constitutionally proper depending on the circumstances), *review granted* (Minn. Ct. App. May 29, 2024), *with* United States v. Smith, 110 F.4th 817, 837 (5th Cir. 2024) (holding that a geofence warrant is akin to a general warrant authorizing a roving generalized search forbidden by the Fourth Amendment), *and* People v. Seymour, 536 P.3d 1260, 1267, 1278 (Colo. 2023) (declining to decide whether individualized probable cause is required for a keyword warrant seeking users who searched for the address of an arson-murder victims' house shortly before the crime because the good faith exception to the exclusionary rule would apply even if probable cause was insufficient).

11. *See* discussion *infra* Part II.A. *See generally* Carpenter v. United States, 138 S. Ct. 2206, 2217 n.3, 2220–21 (2018) (holding that the Fourth Amendment requires police to obtain a warrant before obtaining cell site location information aggregating seven or more days of data).

12. *See* discussion *infra* Part II.A.

13. *See* Ronald J. Allen, *Factual Ambiguity and a Theory of Evidence*, 88 NW. U. L. REV. 604, 632 (1994) ("Any system constructed to make measurements, whether of the past or of something else, may be designed to maximize either accuracy or reliability, or some mix of the two. . . . A judicial system, though, must do more than just reach correct results; it must also be perceived to do so."); Michael S. Pardo, *The Nature and Purpose of Evidence Theory*, 66 VAND. L. REV. 547, 559 (2013) ("The evidentiary proof process may be evaluated based on two considerations: (1) factual accuracy and (2) allocation among the

governmental intrusion, not accuracy in determining guilt or innocence, is a core preoccupation of the modern Fourth Amendment, which indeed is willing to sacrifice accuracy in fact-finding to protect other values.[14] Moreover, evidence law is executed in the trial courts as gatekeeping of the integrity of the fact-finding process, rather than dependent on the highly constrained certiorari determinations of the U.S. Supreme Court, which addresses a miniscule fraction of the cutting-edge controversies in the field—perhaps years to decades later.[15]

For purposes of analysis, this Article focuses on three big data search strategies that are rising in import and stirring conflicts and confusion in the courts. First are keyword warrants requesting that major search companies like Google reveal users who searched keywords connected to a crime.[16] Technology companies like Google amass data on the keywords we search and our online behaviors to target ads and services.[17] These search histories also may build a case of guilt for a crime.[18] Second are

---

parties of the risk of factual errors. Any evidence theory . . . will provide, rely upon, or otherwise presuppose some account of how these considerations relate to the process of proof."); United States v. Havens, 446 U.S. 620, 626 (1980) ("There is no gainsaying that arriving at the truth is a fundamental goal of our legal system."); Tehan v. U.S. ex rel. Shott, 382 U.S. 406, 416 (1966) (explaining that "[t]he basic purpose of a trial is the determination of truth" and safeguards are necessary to avert "the clear danger of convicting the innocent").

14. *See, e.g.*, Tom Stacy & Kim Dayton, *Rethinking Harmless Constitutional Error*, 88 COLUM. L. REV. 79, 89 (1988) ("Fourth amendment [sic] rights, which protect privacy values at the expense of the search for the truth, are examples of such truth-impairing rights.").

15. *See, e.g.*, Barry P. McDonald, Opinion, *This Is the Shadiest Part of the Supreme Court,* N.Y. TIMES (Nov. 3, 2021), https://www.nytimes.com/2021/11/03/opinion/supreme-court-shadow-docket.html (noting that the Supreme Court grants only about one percent of all certiorari petitions).

16. *See* discussion *infra* Part I.A.1.

17. *See* Emilee Rader, *Awareness of Behavioral Tracking and Information Privacy Concern in Facebook and Google* ("Most web pages include code that users cannot see, which collects data necessary for making predictive inferences about what each individual user might want to buy, read, or listen to . . . . [O]nce it has been collected it is not just used to reflect users' own likes and interests back through targeted advertisements."), *in* SOUPS '14: PROCEEDINGS OF THE TENTH USENIX CONFERENCE ON USABLE PRIVACY AND SECURITY 51, 51 (2014); Mary D. Fan, *The Right to Benefit from Big Data as a Public Resource*, 96 N.Y.U. L. REV. 1438, 1440–41 (2021) (describing the numbers of devices creating data for commercial purposes and analysis).

18. *See, e.g.*, People v. Seymour, 536 P.3d 1260, 1268–69 (Colo. 2023) (explaining how keyword warrants led to the discovery of perpetrators of arson-murders).

geofence searches that investigate all cell phones present around the time and place of a crime, essentially turning the convenience of location tracking on our phone and apps into spies and snitches.[19] Third is facial recognition technology (FRT), which searches datasets of images scraped from myriad sources to find a match.[20]

Opponents of big data search strategies couch concerns as privacy claims to fit the Fourth Amendment's modern preoccupation.[21] Yet privacy harms are far less than the costs of inaccuracy resulting in erroneous arrests.[22] At stake is whether the data that companies use every day for profit-making—and even frivolous purposes like how to best sell us cosmetics or sexual performance enhancers—can be used to crack cold cases with unknown perpetrators, where probable cause specific to a particular person does not yet exist.[23] The argument to close access is particularly puzzling because sharing information with third parties usually reduces or eliminates our privacy interests under Fourth Amendment doctrine.[24]

---

19. *See* discussion *infra* Part I.A.2.

20. *See* discussion *infra* Part I.A.3.

21. *E.g.*, United States v. Wright, No. 19-CR-149, 2023 WL 5804161, at *10 (S.D. Ga. Sept. 7, 2023) (describing a constitutional challenge against a geofence warrant under expectations of privacy); United States v. Chatrie, 590 F. Supp. 3d 901, 925–26 (E.D. Va. 2022) (expressing concern about incursions into the privacy of persons by geofence warrants), *aff'd*, 107 F.4th 319 (4th Cir. 2024), *reh'g en banc granted by*, 2024 WL 4648102 (4th Cir. Nov. 1, 2024); Esteban De La Torre, *Digital Dragnets: How the Fourth Amendment Should Be Interpreted and Applied to Geofence Warrants*, 31 S. CAL. INTERDISC. L.J. 329, 345–46 (2022) (arguing geofence warrants violate the privacy expectations of innocent individuals).

22. *See* Slobogin, *supra* note 8, at 960 (arguing that particularized suspicion prevents the identification of suspects and the issuance of geofence warrants generally). Practices, such as geofence warrants, are minimally intrusive. *Id.* at 961 (noting that the "hassle rate" of geofence warrant procedures is minimal because even if numerous anonymized users are revealed in the first step of the process, nobody is even identified, much less "physically hassled," until the list is narrowed to just the most likely suspects).

23. *See* Fan, *supra* note 8, at 894, 930–31 (explaining the potential of "digital search strategies" to help identify "unknown perpetrators").

24. *See* United States v. Miller, 425 U.S. 435, 443 (1976) ("The depositor takes the risk, in revealing his affairs to another, that the information will be conveyed by that person to the Government." (citing United States v. White, 401 U.S. 745, 752 (1971))); Smith v. Maryland, 442 U.S. 735, 737, 742 (1979) (rejecting an argument by petitioner that a pen register's installation (i.e., installation of a device that records dialed phone numbers) was a "search" because all people accept that they give up some privacy to telephone companies).

The more compelling concern is the potential severe harm of over-reliance on potentially inaccurate big data-based search results and digital trails. Consider the case of Harvey Eugene Murphy, Jr., a sixty-one-year-old grandfather: Murphy was in California when two armed robbers struck at a Sunglass Hut in Houston, Texas, holding a manager and associate at gunpoint.[25] The head of loss prevention for the parent company of Sunglass Hut told Houston police investigators that the business's facial recognition technology identified the perpetrator caught on camera.[26] Proprietary facial recognition software combs through vast databanks of images scraped from myriad sources, from Flickr to mugshots to social media, using artificial intelligence to find a match.[27] The software wrongly identified Murphy as the perpetrator of the latest robbery—as well as prior robberies of a Macy's and Sunglass Hut.[28]

Compounding the error, the Houston police showed the manager a picture of Murphy after the loss prevention officers who obtained the facial recognition hit had prepared the witness.[29] Thus tainted, the eyewitness also identified Murphy as the perpetrator.[30] Jailed and denied bond, Murphy alleges that three inmates brutally beat and sexually assaulted him before an attorney won his release by showing he was in California at the time of the robbery.[31] He still has permanent injuries from the jailhouse attack.[32]

---

25. *Murphy* Complaint, *supra* note 1, at 1, 5–6.

26. *Id.* at 4–5.

27. *See* Oliver Hodges, *Facial Recognition Bots Are Scraping Private Data*, SUNDAY TIMES (June 11, 2023), https://www.thetimes.co.uk/article/facial-recognition-bots-scraping-private-data-pfs6p5dlw [https://perma.cc/ZQ2A-8TML] (describing the dangers of facial recognition technology in relation to policing); Kashmir Hill, *The Secretive Company that Might End Privacy as We Know It*, N.Y. TIMES (Jan. 19, 2020), https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html (detailing the use of a new facial recognition AI program by law enforcement—and the concerns that it raises); *see also, e.g.*, Yuni Wen & Matthias Holweg, *A Phenomenological Perspective on AI Ethical Failures: The Case of Facial Recognition Technology*, 39 AI & SOC'Y 1929, 1932–37 (2024) (offering case studies of facial recognition developers scraping images to build datasets to train artificial intelligence algorithms).

28. *Murphy* Complaint, *supra* note 1, at 4–5.

29. *Id.* at 5.

30. *Id.*

31. *Id.* at 6.

32. *Id.*

Consider also the story of Jorge Molina, which has fast become a cautionary tale invoked by opponents of geofence warrants to identify cell phones and their users present at the scene and time of a crime.[33] Molina's story further illuminates that even though opponents of geofence warrants frame their arguments in terms of Fourth Amendment privacy expectations, the heart of the concern is actually inaccuracy. While opponents often frame Molina's experience as one where a geofence warrant led to a wrongful arrest, the details are more complex.[34] Omitted in the accounts by opponents of geofence warrants is that Jorge Molina's abusive stepfather, Marcos Cruz Gaeta, sometimes used Molina's devices, which were logged into Molina's accounts, and Molina's white Honda sedan without permission.[35] Surveillance footage showed that a person in a white Honda sedan drove up to Joseph Knight as he was biking home from work and shot him nine times.[36] When a geofence warrant revealed Molina's old cell phone and white Honda Accord in proximity to the murder of Knight, police arrested Molina.[37] Cruz, with his violent history and practice of taking Molina's cell phone and car, was the likelier perpetrator.[38] Yet investigators with dogged tunnel vision fixated on Molina, ignoring information to the contrary.[39]

---

33. Jorge Molina and Zachary McCoy are frequently used as examples to demonstrate issues with geofence warrants. *See* Amster & Diehl, *supra* note 8, at 397–98 (using the *Molina* and *McCoy* cases as cautionary tales); Edano, *supra* note 8, at 998–99 (noting the *Molina* case); Note, *Geofence Warrants and the Fourth Amendment*, *supra* note 8, at 2508–09 (pointing to *Molina*'s and *McCoy*'s experiences); Slobogin, *supra* note 8, at 961 (noting that *Molina*'s case "is often cited as an example of how geofencing can go awry").

34. *See* sources cited *supra* note 33.

35. *Molina* Complaint, *supra* note 2, at 6, 9 (detailing this information in a complaint); *see also* Meg O'Connor, *Avondale Man Sues After Google Data Leads to Wrongful Arrest for Murder*, PHX. NEW TIMES (Jan. 16, 2020), https://www.phoenixnewtimes.com/news/google-geofence-location-data-avondale-wrongful-arrest-molina-gaeta-11426374 [https://perma.cc/WZ94-D25J] (reporting the same information).

36. Nathan J. Fish, *Man Arrested on Suspicion of Avondale Homicide Through Phone's GPS Location*, ARIZ. CENT. (Dec. 18, 2018), https://www.azcentral.com/story/news/local/southwest-valley-breaking/2018/12/18/man-arrested-avondale-homicide-through-phones-gps-location/2350572002 [https://perma.cc/LL3R-XR7K].

37. *Molina* Complaint, *supra* note 2, at 6–7.

38. *Id*.

39. *See id*. (describing the failure to consider the abusive stepfather); Keith A. Findley & Michael S. Scott, *The Multiple Dimensions of Tunnel Vision in*

The arrest at Molina's workplace shattered the young man's life, resulting in his mugshot and name plastered over the news as a killer, six days of incarceration, and an arrest in his records that jeopardized his career and educational goals.[40] Molina's family and friends worked desperately to prove his innocence, producing text messages, witnesses, and an Uber receipt that proved he was at the movies with friends during the crime and also explaining Cruz's violence.[41] When Molina was finally released, the police only then pursued the far likelier suspect: Cruz.[42]

The geofence data had tortuously led police to the likely perpetrator, Cruz, but at huge collateral cost to Molina, then age twenty-three.[43] His story is a cautionary tale about the perils of failure to further investigate and corroborate leads generated by big data searches. The more complex lesson of the attempt to solve the murder of Knight was that a Geofence warrant can generate leads to identify an unknown perpetrator—but terrible harms can ensue if police leap to arrest based on the lead, without further investigation and corroboration.

A more nuanced approach regulating big data searches and the risk of inaccuracy is needed that goes beyond the perennial Fourth Amendment debates. This Article advances evidentiary safeguards for big data tools with the power to potentially crack cold cases and open leads where perpetrators are unknown.[44] Fourth Amendment law is a blunt and cumbersome instrument to address the complex concerns posed by search strategies drawing on pooled or big data, especially data controlled by private entities.[45] In contrast, evidence regulates what suffices as probable cause for an arrest, or sufficient proof to convict, incentivizing corroboration and further investigation.[46] Evidentiary procedures also permit pretrial evaluations of the reliability of

Criminal Cases, 2006 WIS. L. REV. 291, 292–93 (explaining how tunnel vision can lead to errors in criminal cases).

40. *Molina* Complaint, *supra* note 2, at 7, 10–11.

41. *Id.* at 8–9.

42. Michelle Cruz, *Man Suspected in Avondale Homicide Arrested in California*, ARIZ. REP. (Mar. 13, 2019), https://www.azcentral.com/story/news/local/southwest-valley-breaking/2019/03/13/marcos-cruz-gaeta-suspected-avondale-homicide-arrested-california/3149541002 [https://perma.cc/J3YF-XA2F].

43. *Id.*

44. *See infra* Parts II.B, III.

45. *See infra* Part II.A.

46. *See infra* Part III.A.

big data identification techniques and can be a lever to ensure fair notice and transparency to the defense.[47] Moreover, expert witnesses regulated by evidence law can inform jury deliberations over the weight to give big data search tactics and results, correcting the mystique of seeming machine infallibility.[48]

This Article proceeds in three Parts. Part I explains the rise in law enforcement's use of technologies of identification drawing on pooled or big data, focusing on three major approaches: keyword searches, geofence location-based searches, and facial recognition technology using scraped facial images.[49] There is an important body of literature illuminating the risks of error with even well-settled forensic identification strategies, such as DNA identification or eyewitness identifications, which are the leading causes of wrongful convictions.[50] This Part further explains how technologies of identification drawing on searches of privately held pooled or big data are even more error-fraught and problematic.[51]

Part II argues that evidence law is theoretically and pragmatically suited to address the concerns surrounding the rise of big data suspect identifications.[52] Constitutional criminal procedure, an alluring and popular area for litigation and scholarship, should be complemented by evidence law-based safeguards.[53] This Part also explains how, under the Fourth Amendment's third-party exposure doctrine, the fact that we shared our data with private companies reduces or extinguishes our reasonable expectations of privacy.[54]

---

47. *See infra* Part III.B.

48. *See infra* Part III.C.

49. *See infra* Part I.A.

50. *See infra* Part I.B. *See generally* ERIN E. MURPHY, INSIDE THE CELL: THE DARK SIDE OF FORENSIC DNA 88–143 (2015); Jed S. Rakoff & Elizabeth F. Loftus, *The Intractability of Inaccurate Eyewitness Identification*, DAEDALUS, Fall 2018, at 90, 90–94 [hereinafter MURPHY, INSIDE THE CELL] ("Inaccurate eyewitness testimony is a leading cause of wrongful convictions."); *Eyewitness Misidentification*, INNOCENCE PROJECT, https://innocenceproject.org/eyewitness-misidentification [https://perma.cc/D89U-VKB5] ("Eyewitness misidentification contributes to an overwhelming majority of wrongful convictions that have been overturned by post-conviction DNA testing.").

51. *See infra* Part I.B.

52. *See infra* Part II.A.

53. *See infra* Part II.B.

54. *See infra* Part II.B.

Part III turns to evidence law reforms to address important concerns about suspect identifications using privately-controlled big data.[55] This Part advances three proposals. The first proposal is requiring corroboration before suspect identifications using big data searches can constitute probable cause for an arrest or evidence for conviction.[56] Second, this Article argues for rigorous pre-trial notice, disclosure and reliability testing requirements.[57] Third, before the fruits of big data search results may be introduced, juries must be educated by expert testimony about the risks of error to overcome the mystique of machine infallibility and dangers of over-weighing big data search results.[58]

I. THE RISE OF THE THIRD GENERATION OF TECHNOLOGICAL EVIDENCE

More than a decade and a half ago, Professor Erin Murphy termed forensic technologies such as DNA typing, electronic location tracking, biometric scanning, and data mining the "second-generation" of forensic evidence.[59] To a public dazzled by the power of tools like DNA matching to solve cold cases and exonerate the falsely accused, Professor Murphy sounded caution, recalling the dangers of too readily relying on first-generation techniques like hair, fiber, handwriting, and ballistics analyses.[60] Scandals over laboratory and forensic analyst errors and outright fabrication offered cautionary tales about the veneer of scientific rationality and seeming certitude.[61] While focusing on DNA typing because it was the most developed of the forensic strategies, Professor Murphy presciently predicted:

> It is easy to imagine a future in which evidence culled from cell phones, computers, "EZ Pass" cards, and smart identification cards becomes more ubiquitous, or in which images from a security camera linked to a database facial recognition system are used to convict a host of offenders across a broad spectrum of crimes.[62]

---

55. *See infra* Part III.

56. *See infra* Part III.A.

57. *See infra* Part III.B.

58. *See infra* Part III.C.

59. Erin Murphy, *The New Forensics: Criminal Justice, False Certainty, and the Second Generation of Scientific Evidence*, 95 CALIF. L. REV. 721, 744 (2007) [hereinafter Murphy, *New Forensics*] (coining the term "second-generation forensic evidence").

60. *Id.* at 745–48.

61. MURPHY, INSIDE THE CELL, *supra* note 50, at 49–53.

62. Murphy, *New Forensics*, *supra* note 59, at 728–29.

That future is well upon us, evolving into a third generation of technological evidence drawing on specialized expertise, and privately-collected and -controlled databases and search techniques.

This Part begins by detailing the rise of three major suspect identification strategies drawing on privately-collected and -controlled big data. The first are keyword warrants drawing on search history information held by business that own the major search engines.[63] Second are geofence warrants, drawing on location information collected and stored by Big Tech companies.[64] Third are searches of privately-collected and -controlled facial image databases scraped from public-facing websites and social media.[65] This Part then explains how this third generation of big data-based suspect identification presents important differences from prior generations of scientific-technological evidence, raising three concerns over error and control.[66] The first concern entails private collection and control over the data.[67] The second concern is deviation from the commercial use case of the data.[68] Third are unknown and potentially highly disparate error rates that can vary based on skin tone, race, gender, ethnicity, geographical features, and socioeconomic circumstances.[69]

A. SUSPECT UNKNOWN: CRACKING CASES USING DIGITAL TRAILS

You are walking home from class at dusk. Individuals approach with a gun and demand your backpack, laptop, and cell phone. You hand over the items. But you want your belongings back. How can you figure out where the robbers have absconded with your items?

Kevin Bui, then age sixteen, used the "Find my iPhone" feature to try to locate the robbers who stole his phone and shoes when he attempted to arrange a gun purchase.[70] The attempted

63. *See infra* Part I.A.1.
64. *See infra* Part I.A.2.
65. *See infra* Part I.A.3.
66. *See infra* Part I.B.
67. *See infra* Part I.B.1.
68. *See infra* Part I.B.2.
69. *See infra* Part I.B.3.
70. Natalie Neysa Alund, *He Traced His Stolen iPhone to the Wrong Home and Set It on Fire Killing 5. Now, He Faces Prison.,* (May 22, 2024), https://www

self-help ended tragically because he mistakenly believed he located his phone in a suburban Denver home and set the home afire with his friends in the early morning.[71] Five family members died in the blaze, including young parents Djibril and Adja Diol; their two-year-old daughter, Khadija; Djibril's sister, Hassan; and Hassan's infant daughter, Hawa Beye.[72]

The Senegalese immigrant family's cold case murders went unsolved for six months, terrifying the community with fears that the ghostly masked perpetrators caught on surveillance camera were hate killers.[73] Investigators tried to crack the case as it grew increasingly cold using two techniques drawing on big data searches: geofence and keyword warrants.[74] The keyword warrants unraveled the mystery, leading the police to Bui, who googled the address of the home shortly before the arson.[75] The keyword warrant opened a lead to crack the cold case, revealing that the violence arose from teen and technological error rather than race and immigration status-based animus.[76]

.usatoday.com/story/news/nation/2024/05/21/kevin-bui-senegalese-family-killed-fire-denver/73783060007 [https://perma.cc/DJD2-A8ZC].

71. *Id.*

72. *Id.*; *see also Names, Photos Release of All 5 Victims in Green Valley Ranch Arson*, CBS COLO. (Aug. 13, 2020), https://www.cbsnews.com/colorado/news/victims-green-valley-ranch-arson-djibril-diol [https://perma.cc/G296-GPBC].

73. *Fears Grow for Denver's Senegalese Community as Reward Increases in Green Valley Ranch Fatal Fire Case*, CBS NEWS (Aug. 13, 2020), https://www.cbsnews.com/colorado/news/fears-denver-senegalese-community-reward-increases-green-valley-ranch-fatal-fire [https://perma.cc/4XM4-R3TS] ("It's hard to understand the three people in masks and hooded clothing who investigators believe started the fire at about 2:30 in the morning on Aug. 5. So far there's no suggestion of motive. It is possible that it could be a hate crime in some way, but there's no evidence of that yet."); Darren Whitehead, *Green Valley Ranch Murder Case: Google Evidence Will Be Allowed at Teen's Trial*, 9NEWS (Nov. 16, 2022), https://www.9news.com/article/news/crime/green-valley-ranch-arson-murder/73-b3e6f847-d510-4a2b-bec6-e9351352ffd5 [https://perma.cc/5DG7-SB6Y] (reporting that the "case went cold for months").

74. Motion to Suppress Evidence from a Keyword Warrant & Request for a Veracity Hearing at 3–4, People v. Seymour, No. 21CR20001 (Colo. Dist. Ct. Nov. 16, 2022) (on file with the Minnesota Law Review) [hereinafter Motion to Suppress Evidence (*Seymour*)], *aff'd on other grounds*, 536 P.3d 1260, 1267 (Colo. 2023) ("We discharge the rule to show cause (and thus, essentially affirm the trial court's order), albeit on slightly different grounds.").

75. Response to Motion to Suppress Evidence (*Seymour*), *supra* note 4, at 2–3.

76. *Id.*

The tragedy illustrates the dangers and alluring utility of digital trails and a new generation of technological evidence. Even a teen had the power to try to identify suspects using data, however haphazardly. In the hands of professionals, with the ability to compel production of big data amassed by technology companies or scraped from myriad databases, the power to investigate is even more expansive, powerful, and potentially perilous at large scale.

1. Hunting for Suspicious Search Terms: Keyword Warrants

Keyword warrants, also sometimes referred to as a "reverse-keyword warrant" or "reverse keyword search warrant," draw on the huge pool of data users share with Google.[77] The next time you use Google's search engine, notice whether your account icon shows at the upper right corner of the search page. The dominant search engine provider, Google (with 78.83% of the worldwide search engine market),[78] tracks your search keywords, patterns of behavior, the links you click, and the images and videos you view, among other revealing data.[79] Even if you log out of your Google account in hopes of some semblance of privacy, Google still tracks your location, IP address, the Google services you use, your server information, and the devices you use to access the Web and Google services.[80] Even if you select anonymous browsing, your browsing history and patterns can be linked to your social media profiles, such as your LinkedIn, X (formerly Twitter), Facebook, or Instagram accounts to de-anonymize and identify you.[81]

---

77. *See, e.g.*, People v. Seymour, 536 P.3d 1260, 1267–69 (Colo. 2023) (explaining the process behind a keyword warrant, also referred to by the Justices as a "reverse-keyword warrant" and "reverse keyword searches").

78. *See* Tiago Bianchi, *Market Share of Leading Desktop Search Engines Worldwide from January 2015 to January 2025*, STATISTA (Jan. 23, 2025), https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines [https://perma.cc/L87A-QBEQ] (showing Google's market share by month).

79. Tim Fisher, *How to Stop Google from Tracking Your Searches*, LIFEWIRE (July 16, 2021), https://www.lifewire.com/stop-google-from-tracking-your-searches-4123866 [https://perma.cc/YV5W-JFKC] (explaining what Google tracks).

80. *Id.*

81. *See* Jessica Su et al., *De-Anonymizing Web Browsing Data with Social Networks* ("Can online trackers and network adversaries de-anonymize web browsing data readily available to them? We show—theoretically, via

The vast trove of data that private companies amass on consumer behaviors for profit-making purposes are also alluring for criminal investigations. Keyword warrants, also called reverse keyword search warrants, compel search engine companies to reveal the users who searched specific incriminating keywords within a relevant period for the planning or commission of a crime.[82] Investigators usually direct keyword warrants at three companies that together command nearly the entire search market: Google (83% market share), Microsoft, maker of search engine Bing (9% market share), and Yahoo (3% market share).[83]

In compliance with Google's privacy protocols, the keyword warrants usually entail a three-step process that limits the release of identifying information.[84] In Step One, Google produces a list of anonymized users and devices that have searched the specified suspicious keywords during the relevant time period.[85] This "production version" released to law enforcement gives:

> (1) [T]he date and time of the search, (2) coarse location information inferred from the IP address from which the search was conducted, (3) the Query (search query entered by the user), (4) the Result (the result generated by Google from a user's queried search), (5) the Host (the Google domain name that the user contacted (e.g., google.com and google.fr.)), (6) the Request (the latter part of the URL, following the host, that is associated with the user's search. . . .), (7) a truncated

---

simulation, and through experiments on real user data—that de-identified web browsing histories can be linked to social media profiles using only publicly available data."), *in* WWW '17: PROCEEDINGS OF THE 26TH INTERNATIONAL CONFERENCE ON WORLD WIDE WEB 1261, 1261 (2017).

82. *E.g.*, Affidavit at 4, *In re* Search of Info. & Recs. Associated with Google Searches for Various Search Terms that Are Stored at Premises Controlled by Google, No. 18-mj-00170 (W.D. Tex. Mar. 14, 2018) (filed under seal) (on file with the Minnesota Law Review) [hereinafter Affidavit re Google Searches].

83. Bianchi, *supra* note 78; *see also, e.g.*, Affidavit re Google Searches, *supra* note 82; Affidavit at 4, ¶ 6, *In re* Search of Info. & Recs. Associated with Microsoft Searches for Various Search Terms that Are Stored at Premises Controlled by Microsoft, No. 18-mj-00171 (W.D. Tex. Mar. 14, 2018) (filed under seal) (on file with the Minnesota Law Review) [hereinafter Affidavit re Microsoft Searches]; Affidavit at 1, ¶ 1, *In re* Search of Info. & Recs. Associated with Yahoo Searches for Various Search Terms that Are Stored at Premises Controlled by Oath Holdings, Inc., No. 18-mj-00168 (W.D. Tex. Mar. 14, 2018) (filed under seal) (on file with the Minnesota Law Review) [hereinafter Affidavit re Yahoo Searches].

84. *See infra* Part I.A.2 (explaining the three-step protocol for geofence warrants); *see also, e.g.*, People v. Seymour, 536 P.3d 1260, 1268–69 (Colo. 2023) (describing the three-step warrant process).

85. Declaration of Legal Investigations Support Analyst at 3, ¶¶ 7–8, *Seymour*, 536 P.3d 1260 (No. 2021CR20001).

> Google identifier (known as a GAIA ID), if the search was conducted from an authenticated user's account, or a truncated version of a Browser Cookie ID, if the search was not conducted from an authenticated user's account and (8) the associated user agent string.[86]

In Step Two, investigators review the anonymized list of searches and user locations and winnow down the list of suspects based on contextual factors like whether they were in the same state as the crime site.[87] In Step Three, investigators seek the IP addresses and subscriber information for the narrowed list of the most likely suspects.[88]

The data that keyword warrants can yield are powerfully revealing because of the vast casual surveillance that tech companies maintain on our online behaviors. If a user is logged onto their personal account, the companies have data on associated email addresses, device IP addresses, physical addresses, and other activity conducted by the account holder.[89] Even if a user is not logged onto their personal account, Google, Microsoft, and Yahoo record the IP addresses of devices used to conduct searches.[90] The companies also retain data on patterns of search behavior, including how long we spend online; what sites we use; and times, dates, and places where we log onto associated accounts, such as our Gmail account.[91]

Law enforcement use of keyword warrants is shrouded in secrecy, shielded by protective orders, and sometimes only surfaces because of mistakes.[92] The few known keyword warrants to surface reveal that law enforcement use the strategy as an entryway into an investigation to generate leads to identify unknown perpetrators on the loose, such as the arson-murderers of

---

86. *Id*. at 3, ¶ 7.

87. *Id*. at 3–4, ¶ 8.

88. *Id*. at 4, ¶ 9.

89. Affidavit re Google Searches, *supra* note 82, at 6; Affidavit re Microsoft Searches, *supra* note 83, at 6–7; Affidavit re Yahoo Searches, *supra* note 83, at 6–7.

90. Affidavit re Google Searches, *supra* note 82, at 6; Affidavit re Microsoft Searches, *supra* note 83, at 6; Affidavit re Yahoo Searches, *supra* note 83, at 6.

91. *See, e.g.*, Affidavit, *In re* Search of Info. & Recs. Associated with Google Searches for Various Search Terms that Are Stored at Premises Controlled by Google, No. 18-mj-00189, at 7, ¶¶ 17–18 (W.D. Tex. Mar. 19, 2018) (filed under seal) (on file with the Minnesota Law Review).

92. Jessica Schladebeck, *Feds Issue Secret 'Keyword Warrants' for Google Search History*, GOV'T TECH. (Oct. 7, 2021), https://www.govtech.com/security/feds-issue-secret-keyword-warrants-for-google-search-history [https://perma.cc/39ZA-7YYB].

the Diol family.[93] An example arose during the serial pipe bombings that terrorized Austin, Texas, in March 2018, when packages exploded in the addressee's homes, killing two people and injuring others.[94] Keyword warrants by federal investigators desperate to identify the bomber terrorizing Texas were sealed from public view until the bomber Mark Anthony Conditt, twenty-three, killed himself by detonating a pipe bomb as police closed in on him.[95] After ascertaining that Conditt acted alone and closing the investigation in January 2019, the government moved to unseal the warrants.[96]

A second example of law enforcement use of a keyword warrant arose in the investigation of an unsolved hit-and-run killing. Minnesota physician Cathy Donovan was walking her dogs along Highway 169 in Mille Lacs County when a driver fatally hit her and fled.[97] Dr. Donovan's murder went unsolved for months.[98] Trying to determine the unknown perpetrator, investigators turned to Google seeking users who googled the keywords "Mille Lacs hit and run" or "Highway 169 hit and run" over a four-day period.[99] The tactic was alluring to investigators after keyword warrants helped crack other cases with unknown perpetrators, such as apprehending a serial kidnapper and

---

93. *See, e.g.*, Affidavit re Google Searches, *supra* note 82, at 4 (seeking leads to identify an unknown serial pipe bomber).

94. For information about the bombings, see Emanuella Grinberg & Jason Morris, *Austin Residents Fear that the Explosions May Be Racially Motivated*, CNN (Mar. 15, 2018), https://www.cnn.com/2018/03/15/us/austin-explosion-packages/index.html [https://perma.cc/3X7R-FKJS].

95. For information on the police investigation into Conditt and his eventual suicide, see Jason Hanna et al., *Police: Austin Bomber Left 25-Minute Confession Video on Phone*, CNN (Aug. 31, 2018), https://www.cnn.com/2018/03/21/us/austin-explosions/index.html [https://perma.cc/BX7Q-W94H]; Clint Van Zandt, *What Makes a Serial Bomber Tick?*, ATLANTIC (Mar. 30, 2018), https://www.theatlantic.com/health/archive/2018/03/what-makes-a-serial-bomber-tick/556922.

96. *See* Motion for Limited Unsealing for Multiple Search Warrant Affidavits, *In re* Search of Multiple Sources & Locations Related to the Investigation of Mark Conditt & the Austin Bombings of 2018 at 2, No. 18-mj-00218 (W.D. Tex. Jan. 10 2019) (on file with the Minnesota Law Review).

97. Eric Rasmussen & Ricky Campbell, *High-Profile Hit-and-Run Highlights Controversy over Google Search Warrants*, KSTP (May 14, 2024), https://kstp.com/5-investigates/high-profile-hit-and-run-case-highlights-controversy-over-google-search-warrants [https://perma.cc/L3RE-5VXS].

98. *Id*.

99. *Id*.

rapist in Pennsylvania.[100] Another keyword warrant emerged into public view because of accidental disclosure by the U.S. Department of Justice.[101] The case involved the kidnapping, sexual abuse, and trafficking of a minor.[102] Trying to identify the perpetrator, the keyword warrant sought user information for persons who searched the minor victim's name, her mother's name, or her address during a sixteen-day period surrounding the crime.[103] The outcome of the warrant and investigation are unknown because the government quickly corrected its error and re-sealed the warrant.[104]

The accidentally unsealed warrant in the sensitive investigation shows one rationale for the secrecy: protecting intimate details, such as a minor sexual assault victim's address and name.[105] Keyword searches of a victim's name or address are particularly suspicious where the victim is not a public figure and the address is unlikely to be searched for non-crime-related reasons. Another reason for secrecy is to avoid imperiling an ongoing investigation, such as the efforts by the Federal Bureau of Investigation (FBI) to identify the pipe bomber.[106] If crimes are ongoing, investigators do not want the perpetrators to start

---

100. Eric Rasmussen, *Google 'Keyword Warrant' in Minnesota Now Part of National Debate*, KSTP (June 27, 2024), https://kstp.com/kstp-news/top-news/google-keyword-warrant-in-minnesota-now-part-of-national-privacy-debate [https://perma.cc/95H3-UH6S]; *see also* Commonwealth v. Kurtz, 294 A.3d 509, 524–28 (Pa. Super. Ct. 2023), *appeal docketed*, 306 A.3d 1287 (Pa. 2023) (upholding the use of a reverse keyword warrant in apprehending the rapist of five women).

101. Thomas Brewster, *Exclusive: Government Secretly Orders Google to Identify Anyone Who Searched a Sexual Assault Victim's Name, Address or Telephone Number*, FORBES (Oct. 4, 2021), https://www.forbes.com/sites/thomasbrewster/2021/10/04/google-keyword-warrants-give-us-government-data-on-search-users.

102. *Id.*

103. *Id.*

104. *Id.*

105. *See id.* (noting that when the warrant was unsealed, the minor victim's name, address, and the name of her mother were revealed, jeopardizing their privacy).

106. *See, e.g.*, Brendan J. Lyons, *Search Warrants Are Rarely Unsealed. Here's Why*, TIMES UNION (Aug. 12, 2022), https://www.timesunion.com/state/article/Why-search-warrants-rarely-unsealed-17369233.php [https://perma.cc/L62E-P8ZS] (explaining that warrants are rarely unsealed while an investigation is pending).

concealing or deleting their digital trails or modifying their search terms to avoid specified keyword terms in warrants.[107]

2. From Targeted Advertising to Netting Suspects: Geofence Warrants

In contrast to the secrecy surrounding keyword warrants, revelations by Google document the surging law enforcement use of geofence warrants, rising to more than a quarter of all warrants Google received by 2020.[108] Google is a prime target for geofence warrants because of its market dominance and trove of location data stored in its SensorVault and linked databases whenever people use a Google app or a device running the Android operating system.[109] About 97% of all smartphones in the world either use Google applications, or the Android operating system, or both.[110] Google commanded more than 95% of the mobile search engine market share in the United States as of January 2024.[111] Three of the top five most utilized smartphone apps in the United States are Google apps: Gmail, Google Search, and Google Maps.[112]

Google uses the amassed data to analyze our preferences, detect patterns of behavior, and target advertising, including using geofences to deliver location-based ads, generating billions in

---

107. *See, e.g.*, *In re* Search of Info. that Is Stored at the Premises Controlled by Google LLC, 579 F. Supp. 3d 62, 67 n.1 (D.D.C. 2021) (sealing the warrant application "because the criminal investigation is not public and revealing the existence of the warrant could adversely impact the government's investigation").

108. *Supplemental Information on Geofence Warrants in the United States*, GOOGLE, https://services.google.com/fh/files/misc/supplemental_information_geofence_warrants_united_states.pdf [https://perma.cc/DN52-JQ6S].

109. United States v. Chatrie, 590 F. Supp. 3d 901, 908 (E.D. Va. 2022), *aff'd*, 107 F.4th 319 (4th Cir. 2024), *reh'g en banc granted by*, 2024 WL 4648102 (4th Cir. Nov. 1, 2024); United States v. Smith, No. 21-cr-107, 2023 WL 1930747, at *2 (N.D. Miss. Feb. 10, 2023).

110. *In re* Search of: Info. Stored at Premises Controlled by Google, as Further Described in Attachment A, No. 20 M 297, 2020 WL 5491763, at *4 (N.D. Ill. July 8, 2020).

111. *Mobile Search Engine Market Share United States of America*, STATCOUNTER (Jan. 2024), https://gs.statcounter.com/search-engine-market-share/mobile/united-states-of-america/#monthly-202401-202401-bar [https://perma.cc/P7V6-WU5C].

112. Laura Ceci, *Mobile Audience Reach of Leading Smartphone Apps in the United States in December 2024*, STATISTA (Feb. 10, 2025), https://www.statista.com/statistics/281605/reach-of-leading-us-smartphone-apps [https://perma.cc/C258-VXDS].

revenue.[113] Google claims that it only stores location data when consumers opt in to location history and reporting services.[114] However, independent researchers uncovered that Google apps, such as Google Maps, store location information even if a user does not opt into data-sharing.[115] In 2022, the Arizona Attorney General obtained an $85 million settlement with Google based on a lawsuit alleging that Google deceptively gathered users' location information to sell ads even when the user disabled the Location History setting.[116] According to the Arizona Attorney General, the location data was so lucrative to Google—which derived more than eighty percent of its $161 billion revenue in 2019 from advertising—that the company surreptitiously collected the data through settings such as Web & App Activity even when users deactivated their Location History.[117]

A geofence warrant coopts this gold mine of lucrative data to identify perpetrators of a crime by requesting that Google disclose which devices were at a crime scene or series of connected crimes scenes around the time of the crime.[118] While a geofence for targeted advertising is a perimeter where ads are likely to be effective based on a consumer's location, a geofence for crime investigators is the area where suspects were likely to be located at the time of the crime.[119] Typically deployed to identify

113. Jennifer Valentino-DeVries, *Tracking Phones, Google Is a Dragnet for the Police*, N.Y. TIMES (Apr. 13, 2019), https://www.nytimes.com/interactive/2019/04/13/us/google-location-tracking-police.html. For additional information on the use of geofencing for targeted advertising, see Yi-Jen (Ian) Ho et al., *Distance and Local Competition in Mobile Geofencing*, 31 INFO. SYS. RSCH. 1421, 1421–22 (2020).

114. Brief of Amicus Curiae Google LLC in Support of Neither Party Concerning Defendants' Motion to Suppress Evidence from a "Geofence" General Warrant (ECF No. 29) at 7–8, United States v. Chatrie, 590 F. Supp. 3d 901 (E.D. Va. 2019) (No. 19-cr-00130) [hereinafter Brief of Amicus Curiae (Google)].

115. Ryan Nakashima, *AP Exclusive: Google Tracks Your Movements, Like It or Not*, ASSOCIATED PRESS (Aug. 13, 2018), https://apnews.com/article/north-america-science-technology-business-ap-top-news-828aefab64d4411bac257a07c1af0ecb [https://perma.cc/9WDQ-U6VM].

116. Press Release, Ariz. Att'y Gen., Attorney General Mark Brnovich Achieves Historic $85 Million Settlement with Google (Oct. 4, 2022), https://www.azag.gov/press-release/attorney-general-mark-brnovich-achieves-historic-85-million-settlement-google [https://perma.cc/296A-7BFY].

117. *Id.*

118. *E.g.*, *In re* Search of Info. that Is Stored at the Premises Controlled by Google LLC, 579 F. Supp. 3d 62, 69–70 (D.D.C. 2021).

119. United States v. Asghedom, 992 F. Supp. 2d 1167, 1169–70 (N.D. Ala. 2014); United States v. Rhine, 652 F. Supp. 3d 38, 67 (D.D.C. 2023).

unknown perpetrators of a known crime, geofence warrants particularize "the physical area and the time range in which there is probable cause to believe that criminal activity occurred."[120]

In 2018, the FBI sought its first geofence warrant to identify the perpetrator of serial robberies in Maine.[121] The geofence warrant sought SensorVault data on any phones that location data placed at the scene of nine connected robberies around the thirty-minute interval of the crimes.[122] While Google resisted the geofence warrant, the FBI solved the serial robberies by other forensic and technological means, using DNA, toll pass records, and shoeprints.[123]

Despite this initially rocky start, law enforcement found increasing success with geofence warrants in other investigations to identify perpetrators of crimes such as sexual assaults, murders, break-ins, and more.[124] Federal and state law enforcement geofence warrants directed to Google spiked 1,500% between 2017 and 2018 and leaped another 500% between 2018 and 2019.[125] By 2020, the number of geofence warrants that Google received leaped to 11,554, constituting more than a quarter of all requests received.[126]

The future of Google as a tempting target for geofence warrants for crimes committed in 2024 or beyond is in doubt because of a location data storage change that Google announced in December 2023.[127] Rather than storing user location history in its

120. *Rhine*, 652 F. Supp. 3d at 67; *see also In re* Search of Info. that Is Stored at the Premises Controlled by Google LLC, 579 F. Supp. 3d 62, 69 (D.D.C. 2021) (collecting cases defining "geofence").

121. Donna Lee Elm, *Geofence Warrants: Challenging Digital Dragnets*, CRIM. JUST., Summer 2020, at 7, 8.

122. *Id*.

123. Thomas Brewster, *To Catch a Robber, the FBI Attempted an Unprecedented Grab for Google Location Data*, FORBES (Aug. 15, 2018), https://www.forbes.com/sites/thomasbrewster/2018/08/15/to-catch-a-robber-the-fbi-attempted-an-unprecedented-grab-for-google-location-data.

124. *See, e.g.*, Elm, *supra* note 121, at 8 (describing Raleigh's geofence warrants).

125. Brief of Amicus Curiae (Google), *supra* note 114, at 3.

126. *Supplemental Information on Geofence Warrants in the United States*, *supra* note 108 (navigate to "Download supplemental data as a CSV" at the bottom of the file to see the totals for 2020).

127. Cyrus Farivar & Thomas Brewster, *Google Just Killed Warrants that Give Police Access to Location Data*, FORBES (Dec. 14, 2023), https://www.forbes.com/sites/cyrusfarivar/2023/12/14/google-just-killed-geofence-warrants-police-location-data.

SensorVault, Google announced it would store location history locally on user devices, giving the user more control over whether to keep or delete the data and preventing amassed location search data in centralized Google databases.[128] Unofficially, Google representatives indicate the change was to defeat geofence warrant requests.[129] Other technology companies also store location data, but none have Google's overwhelming dominance on nearly all devices.[130] Geofence warrants that have emerged have mostly targeted Google.[131] Google's SensorVault and connected databases still remain alluring targets for data predating the data storage strategy shift, however, because the policy change does not affect the amassed location data Google has gathered for at least fourteen years.[132]

Google's location data is especially valuable because of its greater precision than cell site location information stored by telecommunications companies, such as cell phone services providers.[133] Cell site location information derives from the connections that your cell phone makes with cell towers to acquire wireless connectivity.[134] From these cell tower connections, a cell phone's location can be triangulated from nearby towers to about three-quarters of a mile.[135] By comparison, Google's location data

128. Marlo McGriff, *Updates to Location History and New Controls Coming Soon to Maps*, GOOGLE: THE KEYWORD (Dec. 12, 2023), https://blog.google/products/maps/updates-to-location-history-and-new-controls-coming-soon-to-maps [https://perma.cc/5UD4-7P5M].

129. Farivar & Brewster, *supra* note 127 ("A current Google employee who was not authorized to speak publicly told *Forbes* that along with the obvious privacy benefits of encrypting location data, Google made the move to explicitly bring an end to such dragnet location searches.").

130. *Id.* ("As *Forbes* has previously reported, practically all geofence warrants are targeted at Google, given its vast amount of search and location data.").

131. *Id.* ("[P]ublic court records nearly always point to a data request from Google over other companies.").

132. Valentino-DeVries, *supra* note 113.

133. *In re* Search Warrant Application for Geofence Location Data Stored at Google Concerning an Arson Investigation, 497 F. Supp. 3d 345, 360 (N.D. Ill. 2020) (noting the higher degree of accuracy than the location data considered in *Carpenter*); Mark Harris, *How a Secret Google Geofence Warrant Helped Catch the Capitol Riot Mob*, WIRED (Sept. 30, 2021), https://www.wired.com/story/capitol-riot-google-geofence-warrant [https://perma.cc/5ERU-6TMN].

134. Carpenter v. United States, 138 S. Ct. 2206, 2211–12 (2018).

135. Harris, *supra* note 133; *see* Brief of *Amici Curiae* Technology Law & Policy Clinic at New York University School of Law & Electronic Frontier

draws on more numerous and precise sources of information (including where your phone accesses Wi-Fi, Bluetooth and GPS connections), placing a connected phone to a few meters or even square feet depending on the location context, such as an urban setting dense with connection points compared to a rural area with more limited connection points.[136]

Through threat of delay and fighting in court, Google induced law enforcement to follow a three-step protocol for geofence warrants to limit the release of identifiable user information.[137] The three-step procedure for geofence warrants is similar to that previously summarized for keyword warrants.[138] In Step One, Google anonymizes the devices present at the time and place specified in the geofence warrant and releases: (1) the latitude, longitude, and timestamps of when and where the device was present, and (2) the source of the location information, such as Wi-Fi, GPS or cell tower data.[139] In Step Two, investigators review the "production list" to winnow the request for user information to the likeliest suspect based on factors such as duration of presence and movements—sometimes requesting more movement information to do the narrowing.[140] In Step Three, investigators request account information for a limited number of devices that circumstances indicate are the likeliest to belong to the perpetrator.[141] The numbers of anonymized devices on the Step One production list and the Step Three release of account

---

Foundation in Support of Defendant-Appellant at 8, United States v. Chatrie, 590 F. Supp. 3d 901 (E.D. Va. 2023) (No. 22-4489) (noting that cell tower data provides less precise location information than geofence data).

136. Criminal Complaint at 2, United States v. Rhine, 652 F. Supp. 3d 38 (D.D.C. 2023) (No. 21-mj-00646) (noting that location is estimated using GPS data, Wi-Fi access points, and Bluetooth beacons); Doug Austin, *Google Geofence Data Identified 5,723 Devices Near January 6th US Capitol Attack: Data Privacy Trends*, E-DISCOVERY TODAY (Dec. 2, 2022), https://ediscoverytoday.com/2022/12/02/google-geofence-data-identified-5723-devices-near-january-6th-us-capitol-attack-data-privacy-trends [https://perma.cc/B37L-AFKN] (discussing the same); Jennifer Lynch, *Google's Sensorvault Can Tell Police Where You've Been*, ELEC. FRONTIER FOUND. (Apr. 18, 2019), https://www.eff.org/deeplinks/2019/04/googles-sensorvault-can-tell-police-where-youve-been [https://perma.cc/9597-PKYW].

137. Response to Rule 17 Subpoena attach. B, at 2, United States v. Chatrie, 590 F. Supp. 3d 901 (E.D. Va. 2022) (No. 19-cr-00130).

138. See *supra* text accompanying notes 84–88.

139. *Id.* at 3, ¶ 8.

140. *Id.* at 4, ¶ 10.

141. *Id.* ¶ 12.

information vary greatly between geofence warrants, depending on the nature of the crime, the setting and time of day when the offense occurred, and the relevant window of time.[142]

For example, in *Chatrie*, there was an investigation of a bank robbery around closing time (4:52 p.m.) in an urban Virginia area, where agents obtained a geofence warrant for the 150-meter radius of the targeted bank for the one-hour period surrounding the offense.[143] The total area encompassed was longer than three football fields and included a nearby church area that surveillance cameras showed the robber crossed.[144] In Step One, Google's production list returned location history for nineteen anonymized devices in the area.[145] Confidence intervals for each device varied in area size depending on the source of the location data and the setting, with the largest confidence interval being 387 meters, or longer than four football fields.[146] After negotiations with Google and visualizing the anonymized devices on a map of the crime scene, investigators winnowed down the list to just three devices for account disclosure at Step Three, ultimately leading investigators to the defendant.[147]

A notably expansive and successful use of a geofence warrant to unmask suspects came in the investigation of the January 6, 2021, insurrection and attack on the U.S. Capitol to disrupt the peaceful transfer of presidential power.[148] Investigators sought device information for persons present within a geofence area roughly following "the contours of the Capitol building itself, excluding most of the plazas and lawns on both sides of the building and the abutting streets."[149] To distinguish rioters from employees at the Capitol, the geofence sought de-identified data for persons present between 2:00 and 6:30 p.m., the time of the

142. *Compare, e.g.*, United States v. Rhine, 652 F. Supp. 3d 38, 86 (D.D.C. 2023) (finding 5,723 anonymized devices at Step One and 1,535 users at step three in the investigation of the insurrection at the U.S. Capitol), *with* United States v. Chatrie, 590 F. Supp. 3d 901, 920–21 (E.D. Va. 2022) (finding nineteen anonymized devices at Step One and three devices at Step Three in investigation of a bank robbery), *aff'd*, 107 F.4th 319 (4th Cir. 2024), *reh'g en banc granted by*, 2024 WL 4648102 (4th Cir. Nov. 1, 2024).

143. *Chatrie*, 590 F. Supp. 3d at 918.

144. *Id.*

145. *Id.* at 920.

146. *Id.* at 922.

147. *Id.* at 921, 924.

148. United States v. Rhine, 652 F. Supp. 3d 38, 46 (D.D.C. 2023).

149. *Id.* at 68.

riots, and control lists of devices present around 12:00 to 12:15 p.m., and 9:00 to 9:15 p.m., which would indicate likely employees rather than rioters.[150] Deleting the likely employees from the list of devices present during the time of the riot yielded a production list of 5,518 anonymized devices at Step One.[151] Focusing on devices that location history indicated were inside the Capitol during the attack, and where the margin of error was entirely within the geofence parameters, the investigators winnowed down the list to the disclosure of account details for 1,498 devices at Step Three.[152] The government also obtained account information for an additional thirty-seven devices that location history obtained on January 6 showed were inside the geofence (bringing the total to 1,535), but by January 7, the history had been deleted, suggesting an attempt to conceal participation in the insurrection.[153]

Here is a twist in the tale of two geofence warrants of different breadths and numbers of affected devices—district judges ruled in different directions. A district judge invalidated one of the warrants while another judge upheld the other warrant.[154] If you were to guess which warrant was invalidated, you might think it was the massive Capitol-area geofence warrant netting 1,535 account user details at Step Three, rather than the modest geofence warrant in *Chatrie* that obtained only three account user details at the same stage.[155] The outcomes are the opposite of such an expectation. Two district judges upheld the U.S. Capitol geofence warrant netting a massive 5,518 anonymized devices and 1,535 user account details.[156] The geofence that a district judge ruled invalid, albeit exempt from exclusion under the good faith exception to the exclusionary rule, was the far more modest bank robbery warrant netting only three user account

150. *Id.* at 68–69.

151. *Id.* at 69.

152. *Id.*

153. *Id.* at 70.

154. *Compare id.* at 82–89 (upholding a geofence warrant), *and* Transcript of Motion Hearing at 27–28, United States v. Cruz, No. 22-cr-00064 (D.D.C. Mar. 3, 2024) (same), *with* United States v. Chatrie, 590 F. Supp. 3d 901, 934–36 (E.D. Va. 2022) (holding that the geofence warrant in question did not satisfy the Fourth Amendment standard for particularity), *aff'd*, 107 F.4th 319 (4th Cir. 2024), *reh'g en banc granted by*, 2024 WL 4648102 (4th Cir. Nov. 1, 2024).

155. *See Chatrie*, 590 F. Supp. 3d, at 921, 924.

156. *Rhine*, 652 F. Supp. 3d at 82–89; Transcript of Motion Hearing at 27–28, *Cruz*, No. 22-cr-00064.

details—one of them being the defendant.[157] A panel of the Fourth Circuit reversed the *Chatrie* district judge and upheld the geofence warrant—but a majority of the active Fourth Circuit granted rehearing en banc.[158] The twists illustrate the mixed and murky legal terrain on geofence warrants.

3. Scraping Images for a Suspect: Facial Recognition Technology

Facial recognition technology is a third controversial, powerful, and lucrative method to identify potential suspects, drawing on vast privately-controlled databanks and proprietary artificial intelligence-powered search algorithms.[159] While some law enforcement agencies have used facial recognition technology to identify suspects for more than twenty years, drawing on state databases such as driver's license records,[160] modern facial recognition technology exists at a massively larger scale, powered by profit-making businesses.[161] Because of controversies and concern over mass surveillance of protesters, racially disparate risks of error, and racial profiling, large companies like Amazon and Microsoft have stopped selling facial recognition technology to law enforcement and businesses.[162] The motherlode of location and search data, Google, refrained from facial

157. *Chatrie*, 590 F. Supp. 3d at 934–36.

158. United States v. Chatrie, 107 F.4th 319, 330–32 (4th Cir. 2024), *reh'g en banc granted by*, 2024 WL 4648102 (Nov. 1, 2024).

159. *See, e.g.*, Brenda M. Simon & Ted Sichelman, *Data-Generating Patents*, 111 NW. U. L. REV. 377, 402 (2017) (discussing patent and trade secret protection for facial recognition methods); Complaint at 3–4, *In re* Everalbum, Inc., No. C-4743 (FTC May 6, 2021), https://www.ftc.gov/system/files/documents/cases/1923172_-_everalbum_complaint_final.pdf [https://perma.cc/D8WY-HVS7] (discussing the harvesting of photos to create a proprietary dataset of faceprints for sale to customers).

160. Jennifer Valentino-DeVries, *How the Police Use Facial Recognition, and Where It Falls Short*, N.Y. TIMES (Jan. 12, 2020), https://www.nytimes.com/2020/01/12/technology/facial-recognition-police.html (describing the facial recognition system in Florida).

161. *See, e.g.*, Kashmir Hill, *The Secretive Company that Might End Privacy as We Know It*, N.Y. TIMES (Jan. 18, 2020), https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html (describing Clearview AI's business and far larger database compared to law enforcement organizations).

162. Jay Greene, *Microsoft Won't Sell Police Its Facial-Recognition Technology, Following Similar Moves by Amazon and IBM*, WASH. POST (June 11, 2020), https://www.washingtonpost.com/technology/2020/06/11/microsoft-facial-recognition.

recognition technology out of concern over misuse.[163] Smaller privately held companies have entered the lucrative market, each with its propriety blend of artificial intelligence-powered facial recognition products and databases of faceprints.[164]

The largest database of facial images for mining by artificial intelligence search algorithms is not owned or controlled by the government.[165] Rather, a small privately-owned company amassed the collection of images of people by scraping popular web sites and apps such as Facebook, YouTube, and Venmo.[166] The technology company that exemplifies the secrecy, controversy, and lucrative nature of selling facial recognition technology, Clearview AI, launched in 2017 with an artificial intelligence-powered facial recognition tool and a proprietary database of faceprints.[167] In early 2022, the company reported to investors that it is on target to amass 100 billion faceprints of people within a year so that "almost everyone in the world will be identifiable."[168] By 2023, the number of faceprints in the Clearview AI databases reached thirty billion faces—a breathtakingly rapid growth fueled by the stream of images Internet users put online and onto the Internet of Things via networked devices.[169]

Clearview AI's proprietary search technology uses neural networks to transform facial images into mathematical terms, called vectors, calculated by using facial geometry such as the

---

163. Bianca Bosker, *Facial Recognition: The One Technology Google Is Holding Back*, HUFFINGTON POST (June 1, 2011), https://www.huffpost.com/entry/facial-recognition-google_n_869583 [https://perma.cc/3XJE-8NKL].

164. *E.g.*, Drew Harwell, *Unproven Facial-Recognition Companies Target Schools, Promising an End to Shootings*, WASH. POST (June 7, 2018), https://www.washingtonpost.com/business/economy/unproven-facial-recognition-companies-target-schools-promising-an-end-to-shootings/2018/06/07/1e9e6d52-68db-11e8-9e38-24e693b38637_story.html.

165. *See* Hill, *supra* note 161.

166. *Id.*

167. *See* Drew Harwell, *Facial Recognition Firm Clearview AI Tells Investors It's Seeking Massive Expansion Beyond Law Enforcement*, WASH. POST (Feb. 17, 2022), https://www.washingtonpost.com/technology/2022/02/16/clearview-expansion-facial-recognition (detailing the history of Clearview AI).

168. *Id.* For more information on Clearview's technology, see Terence Liu, *How We Store and Search 30 Billion Faces*, CLEARVIEW AI (Apr. 18, 2023), https://www.clearview.ai/post/how-we-store-and-search-30-billion-faces [https://perma.cc/8K3L-VALE]. *Cf. Request a Trial*, CLEARVIEW AI, https://engage.clearview.ai/request-a-trial-1 [https://perma.cc/TCM4-6ENQ] (offering free trials to law enforcement, government, and military personnel).

169. Liu, *supra* note 168.

distance between a person's eyes.[170] A neural network is an artificial intelligence technique inspired by how our brains and nervous system work, that uses deep learning, trained on datasets, to process information.[171] Once transformed mathematically into a vector, the facial image to be identified (called a probe) is compared to other images in the databases for potential matches ranked by their similarity.[172]

Clearview AI sells its facial recognition technology to law enforcement agencies and the military and even offers agencies a free trial of its software.[173] Clearview AI formerly also offered its facial recognition technology to private companies, such as Macy's.[174] After the *New York Times* revealed the vast faceprint database of the small company generated by image-scraping, however, lawsuits launched, including cases brought under the Illinois Biometric Information Privacy Act.[175] The Illinois law requires that any business collecting biometric data such as faceprints, iris scans, or fingerprints, from Illinois residents obtain written informed consent and bars profiting from biometric data.[176] The Act also created a private right of action to sue to enforce the protections.[177] The lawsuits alleged that Clearview AI collected facial images without consent in violation of state law.[178] Pursuant to a settlement with the ACLU and other public interest groups in May 2022, Clearview AI agreed to cease

170. *Id.*

171. Terrence J. Sejnowski, *The Unreasonable Effectiveness of Deep Learning in Artificial Intelligence*, 117 PROC. NAT'L ACADEMY SCIS. 30033, 30036–37 (2020).

172. Liu, *supra* note 168.

173. Harwell, *supra* note 167; *Request a Trial*, *supra* note 168.

174. Class Action Complaint ¶¶ 3–6, Carmine v. Macy's Retail Holdings, No. 20-cv-4589 (N.D. Ill. Aug. 5, 2020); Madeline Mitchell, *Macy's Faces Class Action Lawsuit for Use of Facial Recognition Software Clearview AI*, CINCINNATI ENQUIRER (Aug. 7, 2020), https://www.cincinnati.com/story/news/2020/08/07/macys-faces-class-action-lawsuit-use-facial-recognition-software-clearview-ai/3315099001 [https://perma.cc/4GZS-DLQC].

175. Class Action Complaint, *supra* note 174; *see also* 740 ILL. COMP. STAT. 14/1 (2024) (noting the name of the act).

176. *Id.* § 14/15.

177. *Id.* § 14/20.

178. Class Action Complaint, *supra* note 174.

selling to private businesses and persons and instead focus on selling its services to law enforcement and military agencies.[179]

Other private companies are filling the demand from businesses for facial recognition technology, such as retailers struggling with merchandise theft from organized rings that sell the stolen products online.[180] A privately held company, FaceFirst, counts as its clients a quarter of North America's largest retailers and also markets its services to other businesses, such as casinos and event security.[181] Companies such as FaceFirst, Suspect Technologies, and Face-Six also are selling facial recognition products to the educational market, spanning daycares through high school and college, including campus security.[182] The market demand for facial recognition technologies is so alluring that a major company, IBM, reentered the facial recognition market after withdrawing.[183]

While search algorithms vary, the facial recognition technologies deploy a one-to-many search that compares the target image to be identified with a database of faces scraped from the web.[184] In contrast, the facial recognition technology on features (such as the Face ID on an iPhone or a banking app biometric unlock) draws on "one-to-one" facial matching that compares a face attempting to unlock an app or device with the stored

---

179. *Settlement Agreement & Release*, AM. C.L. UNION 1–2 (May 4, 2022), https://www.aclu.org/sites/default/files/field_document/exhibit_2_signed_settlement_agreement.pdf [https://perma.cc/X565-6ZVV].

180. Lauren Debter, *Retailers Quietly Deploying Controversial Technology to Combat Crime Spree*, FORBES (Jan. 31, 2022), https://www.forbes.com/sites/laurendebter/2022/01/31/retailers-quietly-deploying-controversial-technology-to-combat-crime-spree/?sh=1eb166807689.

181. *Id. See generally* FACEFIRST, https://www.facefirst.com [https://perma.cc/GGS8-S7LU].

182. *See, e.g.*, Shuran Zhao, *Facial Recognition in Educational Context*, 586 ADVANCES SOC. SCI. EDUC. & HUMANS. RSCH. 10, 10–11 (2021) (discussing the use of facial recognition at schools and universities).

183. Mark Wilding, *IBM Promised to Back Off Facial Recognition—Then It Signed a $69.8 Million Contract to Provide It*, VERGE (Aug. 31, 2023), https://www.theverge.com/2023/8/31/23852955/ibm-uk-government-contract-biometric-facial-recognition [https://perma.cc/8UXC-AYDQ].

184. *Facial Recognition Technology (FRT)*, NAT'L INST. OF STANDARDS & TECH. (Feb. 6, 2020) [hereinafter *Testimony of Romine*], https://www.nist.gov/speech-testimony/facial-recognition-technology-frt-0 [https://perma.cc/R247-H3KT] (recording testimony of Dr. Chuck Romine, Director of the Information Technology Laboratory (ITL) at the Department of Commerce's National Institute of Standards and Technology (NIST)).

verified user.[185] Early commercial facial recognition algorithms were trained on datasets with predominantly "lighter-skinned" white faces.[186] The result was racially disparate accuracy rates, with greater accuracy for white faces and worse rates of both false positives (a false match) and false negatives (failure to identify match) for Asian, Black, and Latinx faces.[187] Companies have tried to address these defects with training on more diverse datasets culled from the ever-growing number of faces that can be scraped from the Internet and Internet of Things, resulting in substantially improved accuracy.[188] Racially-disparate error rates remain, however, known as "algorithmic bias," a concept that is opaque to members of the public even as more awareness has grown in academia.[189] The lived experience of these errors is severe, as discussed further in Part I.B.3.[190]

B. NEW CHALLENGES POSED BY THE THIRD GENERATION OF TECHNOLOGICAL EVIDENCE

The third generation of suspect identification strategies drawing on commercial big data poses a further set of heightened concerns than did the first or second generations of forensic identification like fingerprints or DNA typing.[191] This section delves into each of the three clusters of factors. First is the creation and control of databanks by private commercial entities, which enjoy trade secret protection and property interests in data and

185. Lindsey Barrett, *Ban Facial Recognition Technologies for Children—and for Everyone Else,* 26 B.U. J. SCI. & TECH. L. 223, 232 (2020).

186. Kimmo Kärkkäinen & Jungseock Joo, *FairFace: Face Attribute Dataset for Balanced Race, Gender, and Age for Bias Measurement and Mitigation*, 2021 IEEE WINTER CONF. ON APPLICATIONS COMPUT. VISION, 1547, 1548.

187. K.S. Krishnapriya et al., *Issues Related to Face Recognition Accuracy Varying Based on Race and Skin Tone*, 1 IEEE TRANSACTIONS ON TECH. & SOC'Y 8, 8–9 (2020).

188. *See, e.g.*, *Testimony of Romine*, *supra* note 184 (noting "massive" improvements in facial recognition search algorithm accuracy between 2013 and 2018).

189. Daniella Raz et al., *Face Mis-ID: An Interactive Pedagogical Tool Demonstrating Disparate Accuracy Rates in Facial Recognition*, 2021 PROC. 2021 AAAI/ACM CONF. ON AI ETHICS & SOC'Y 895, 896.

190. *See supra* Part I.B.3 (discussing racially-disparate error rates).

191. *See* Murphy, *New Forensics*, *supra* note 59, at 726–56 (defining the first and second generations of forensic evidence and the flaws with each, with a focus on DNA typing as the archetypal second-generation form of forensic evidence).

algorithms.[192] Second is the sharp deviation from the original commercial use case when the data is deployed for suspect identifications.[193] A third major factor is highly variable and potentially sharply disparate burdens of error by race, ethnicity, geography, and socioeconomic status.[194]

Discussing the second generation of forensic identification strategies, with a focus on DNA typing, Professor Murphy noted five major differences from first-generation traditional techniques such as: bite, hair, or fiber analyses; handwriting and voice exemplars; and fingerprints.[195] First was application to a wide variety and proportion of cases rather than a more limited subset, such as those involving handwritten materials.[196] Second was the reliance on more rigorous, specialized knowledge that leads to expressions of higher confidence levels and claims of certainty.[197] Third is the costly nature of testing, leading to barriers to verification.[198] Fourth is the incorporation of proprietary technologies, such as the chemical sequences used to conduct DNA analyses.[199] Fifth is the reliance on large computerized government databases like the FBI's Combined DNA Index System (CODIS), a national DNA database that poses privacy concerns.[200]

The third generation of privately-controlled big data suspect identification strategies also apply to a wide swathe of cases, draw on specialized knowledge with the mystique of seeming

192. *See* discussion *supra* Part I.B.1.

193. *See* discussion *supra* Part I.B.2 (discussing the deviation from the commercial use case).

194. *See* discussion *supra* Part I.B.3 (describing unknown, highly variable, and racially disparate risks of error).

195. *See* Murphy, *New Forensics*, *supra* note 59, at 726–31 (comparing first-generation and second-generation forensic identification strategies).

196. *Id*. at 728–29.

197. *See id*. at 729 (noting that second-generation strategies rely on highly specialized knowledge that is consequently purported to be, and viewed by many lay people, as highly probative).

198. *See id*. (explaining that the costliness of second-generation tools means that independent analyses require significant capital expenditure).

199. *See id*. at 729–30 ("DNA typing has . . . weathered a series of challenges related to the reluctance of private companies to divulge claimed proprietary secrets, such as the chemical sequences used to conduct the analysis.").

200. *See id*. (discussing the reliance of second-generation on computerized databases); *id*. at 739 (describing the FBI's Combined DNA Index System); *id.* at 749 (asserting that the reliance of second-generation tools on government-controlled databases raises privacy concerns).

certainty, have barriers to verification, and rely on proprietary technologies.[201] These challenges are amplified by private control of the data, deviation from the commercial use cases, and hard-to-quantify and potentially disparate error rates by race, ethnicity, social and geographical differences.[202]

1. Private Control of Big Databanks and Search Capacities

When law enforcement officers place a GPS tracking device on a vehicle or gather DNA from a crime scene, the collection and control of the data is concentrated in government actors, subject to laws, policies, and procedures for evidence-gathering.[203] In contrast, when Google amasses search and location history, or privately-controlled companies sell facial recognition technology, the data collection is guided by a profit-making imperative.[204] Businesses are not subject to Fourth Amendment regulation for searches and seizures, which is reserved for government actors, not private actors.[205] Businesses also are not democratically accountable to the electorate.[206] In contrast, police chiefs and departments are answerable to elected officials, such as a

201. *See* discussion *infra* Part I.A. (discussing features of third-generation strategies).

202. *See* Parts I.B.1–B.3 (identifying private control, deviation from commercial use cases, and disparate error rates as problems in third-generation strategies).

203. *See, e.g.*, United States v. Jones, 565 U.S. 400, 404–06 (2012) (applying the Fourth Amendment trespass doctrine to the placement of a GPS tracker on a suspect's vehicle); 34 U.S.C. § 40702 (prescribing protocols for the collection of DNA samples from persons arrested, facing charges, or convicted); *DNA Evidence: Basics of Identifying, Gathering and Transporting*, NAT'L INST. OF JUST. (Aug. 8, 2012), https://nij.ojp.gov/topics/articles/dna-evidence-basics-identifying-gathering-and-transporting [https://perma.cc/98DZ-RRQW] (describing procedures for the collection of DNA evidence).

204. *See infra* Part I.A. (discussing the use of privately-collected data in criminal investigations).

205. *See, e.g.*, Skinner v. Ry. Lab. Execs.' Ass'n, 489 U.S. 602, 614 (1989) ("Although the Fourth Amendment does not apply to a search or seizure, even an arbitrary one, effected by a private party on his own initiative, the Amendment protects against such intrusions if the private party acted as an instrument or agent of the Government.").

206. *See*, *e.g.*, Sarah C. Haan, *Civil Rights and Shareholder Activism:* SEC v. Medical Committee for Human Rights, 76 WASH. & LEE L. REV. 1167, 1218 (2019) (distinguishing shareholder governance from democratic accountability).

municipal mayor or city council, and subject to constitutional constraints on power.[207]

Unlike private businesses, law enforcement agencies increasingly post departmental manuals containing policies and procedures for the exercise of power—sometimes pursuant to legislative mandate.[208] Police departments also are increasingly subject to community comment procedures for policies and community oversight boards, especially in the aftermath of protests and calls for reform following the murder of George Floyd.[209] Some legislatures also have mandated that law enforcement agencies promulgate policies on the use of technologies in policing, such as police-worn body cameras.[210] The comparison of the relatively greater constraints on law enforcement investigative practices is not meant to be sanguine about the notorious opacity and oversight challenges in policing, despite decades of attempted reforms.[211] Rather, understanding the comparatively greater constraints on law enforcement agencies compared to businesses in gathering and using data illuminates the grave concerns with using privately collected and controlled data as a basis for identifying suspects and justifying stops, arrests, and convictions.

---

207. *See, e.g.*, Barry Friedman & Maria Ponomarenko, *Democratic Policing*, 90 N.Y.U. L. REV. 1827, 1831 (2015) (noting that "police chiefs typically serve at the pleasure of the mayor, police commission, or city council, and sheriffs are directly elected by the people," though "these oversight mechanisms are no substitute for . . . legislative authorization and public rules").

208. *See, e.g., Administrative Guide,* N.Y.C. POLICE DEP'T (Apr. 4, 2019), https://www.nyc.gov/site/nypd/about/about-nypd/manual.page [https://perma.cc/LPG7-TJEC] ("New York City Local Law No. 129 of 2016, mandates the New York City Police Department to publish the Patrol Guide online for the public to view."); C.J. Ciaramella, *Police Manuals*, NEOCITIES, https://policemanuals.neocities.org [https://perma.cc/2P2Q-SLJY] (containing links to police department manuals for the thirty-eight largest cities in the United States).

209. *See* Sharon R. Fairley, *Survey Says: The Development of Civilian Oversight of Law Enforcement Skyrockets in the Wake of George Floyd's Killing*, 31 S. CAL. REV. L. & SOC. JUST. 283, 295–316 (2022) (discussing the rise of community oversight and civilian review boards after the murder of George Floyd).

210. *See, e.g.*, 50 ILL. COMP. STAT. 706/10-20 (2024) (mandating written policies governing body-worn cameras).

211. *See, e.g.*, Nirej Sekhon, *Dangerous Warrants*, 93 WASH. L. REV. 967, 1013 (2018) ("Police departments are notoriously opaque with regard to policymaking and implementation.").

Intellectual property laws add an additional layer of opacity to private collection and control of data.[212] Copyright and trade secret protections for commercial databases and proprietary algorithms create protections against independent testing, verification, and replication of results.[213] Courts have accorded copyright protection to commercial compilations of consumer data based on the compilation, selection, and arrangement of information into a protectable work.[214] Private firms have powerful incentives to keep machine learning algorithms and training data secret to retain competitive commercial advantages and prevent reverse engineering.[215] Trade secret protections require the maintenance of secrecy to keep data compilations and algorithms proprietary, incentivizing opacity.[216] Even when private companies contract with public entities, such as law enforcement agencies, contractual provisions can require secrecy and non-disclosure to protect proprietary interests.[217]

---

212. *See, e.g.*, Simon & Sichelman, *supra* note 159, at 401–02 (discussing intellectual property protections for privately-collected data and algorithms); Rebecca Wexler, *Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System,* 70 STAN. L. REV. 1343, 1358–64 (2018) (chronicling trade secret protections for forensic technology).

213. *See, e.g.*, Fan, *supra* note 17, at 1466–68 (detailing intellectual property law barriers to accessing commercial data compilations).

214. *E.g.*, Experian Info. Sols., Inc. v. Nationwide Mktg. Servs., Inc., 893 F.3d 1176, 1179–80, 1186 (9th Cir. 2018) (extending copyright protection to company Experian's consumer database); *In re* Nw. Airlines Priv. Litig., No. Civ. 04-126, 2004 WL 1278459, at *4 (D. Minn. June 6, 2004) ("[W]hen that information was compiled and combined with other information to form a [passenger name record (PNR)], the PNR itself became Northwest's property."); Mason v. Montgomery Data, Inc., 967 F.2d 135, 136, 140–42 (5th Cir. 1992) (recognizing copyright protection for the compilation and arrangement of real estate data visualized on a map).

215. *See* W. Nicholson Price II & Arti K. Rai, *Clearing Opacity Through Machine Learning*, 106 IOWA L. REV. 775, 791–92 (2021) ("The central obstacle to reducing tool opacity—and to reproducibility by competitors—is secrecy with respect to the learning algorithm, training data, training process, and associated parameters. . . . [C]ompetitive advantage can be an important incentive.").

216. *See* Ruckelshaus v. Monsanto Co., 467 U.S. 986, 1011 (1984) ("With respect to a trade secret, the right to exclude others is central to the very definition of the property interest. Once the data that constitute a trade secret are disclosed to others, or others are allowed to use those data, the holder of the trade secret has lost his property interest in the data.").

217. *See, e.g.*, Elizabeth A. Rowe & Nyja Prior*, Procuring Algorithmic Transparency*, 74 ALA. L. REV. 303, 322–23 (2022) (discussing "stringent" nondisclosure provisions in contracts between states and private software companies).

Trade secret-related barriers to transparency also potentially exist with other forensic techniques, such as DNA analyses.[218] As Rebecca Wexler chronicles, state courts in California, New York, Ohio, Pennsylvania, and Washington have denied defense requests for the proprietary source code for probabilistic genotyping software, such as TrueAllele (owned by Cybergenetics) or New York's Forensic Statistical Tool, which computes the likelihood that the suspect contributed a DNA sample compared to the likelihood that someone else contributed the sample.[219] Other courts, however, have ordered disclosure of the source code of the software under protective order, balancing defense interests with proprietary interests.[220]

The opacity induced by intellectual property protections and commercial secrecy incentives is greater in degree and complexity for geofence, keyword, and facial recognition searches. Proprietary analytical software is used in a step of DNA analyses that can also be explicated by experts or even disclosed under protective order.[221] Other crucial steps, including the DNA collection, storage, processing, and chain-of-custody matters, are established via government witnesses and subject to challenge if there are deviations from standard procedures or best practices for evidence handling.[222] In contrast, with big data suspect identification techniques, the compilation and search processes are all controlled by private entities for profit-making rather than

218. *See* Wexler, *supra* note 212, at 1362 n.80 (listing cases where defendants have been denied access, on trade-secret grounds, to the technologies used to analyze their DNA).

219. *Id.*; *see also* People v. Wakefield, 195 N.E.3d 19, 28 (N.Y. 2022); People v. Williams, 147 N.E.3d 1131, 1135 (N.Y. 2020).

220. *See, e.g.*, State v. Pickett, 246 A.3d 279, 284, 301 (N.J. Super. Ct. App. Div. 2021) (ordering disclosure under protective order and noting other similar decisions).

221. *Id.* at 298–300 (discussing disclosure under protective order); *cf.* Williams v. Illinois, 567 U.S. 50, 57–58 (2012) (holding that DNA evidence may be subject to cross-examination).

222. *See, e.g.*, Susan Ballou et al., *The Biological Evidence Preservation Handbook: Best Practices for Evidence Handlers*, NAT'L INST. OF JUST. 9–35 (2013), https://nvlpubs.nist.gov/nistpubs/ir/2013/NIST.IR.7928.pdf [https://perma.cc/8TYL-E33K] (instructing law enforcement on best practices for collecting, storing, and maintaining chain-of-custody for biological evidence such as DNA, hair, and fibers); Hill v. State, 269 So. 3d 1, 9 (Miss. 2018) (describing how defense counsel questioned witnesses about DNA storage and chain of custody issues); State v. Bailey, 677 N.W.2d 380, 394 (Minn. 2004) (discussing how the government presented witness testimony about the chain of custody and storage of DNA).

evidentiary purposes.[223] Disclosure of vast proprietary databases outside the custody and control of the government and of the algorithms used to manage and search them is far more complex and impracticable compared to disclosing the source code of a probabilistic genotyping software.[224]

Moreover, government actors subject to constitutional, statutory, and standard-setting regulation maintain national fingerprint and DNA databases.[225] In contrast, the commercial big data used in the third generation of suspect identification techniques are collected, controlled, and owned by private actors.[226] This makes attempts at oversight even more challenging: whereas the government provides for quality assurance testing for forensic DNA testing laboratories and access to CODIS, such access is impeded for privately collected and owned data.[227]

2. Deviating from the Use Cases of the Commercial Data

Another shared feature of the third generation of big data search strategies is the co-optation of private data and algorithms from their original use case when deployed for suspect

223. *See supra* Parts I.A.1–A.3 (discussing data collection and suspect-identification techniques conducted by private entities).

224. *Cf., e.g.*, Government's Response in Opposition to Defendant's Motion for Discovery of Sensorvault Data at *4, United States v. Chatrie, 590 F. Supp. 3d 901 (E.D. Va. Nov. 19, 2019) (No. 19-cr-00130), 2019 WL 7660963 at *4 [hereinafter Gov't Response to Discovery (*Chatrie*)] (demonstrating the difficulty of DNA evidentiary disclosure where Google's Sensorvault data, its parameters, and the qualifications of Google employees were outside of the custody and control of the government).

225. *See, e.g.*, 34 U.S.C. § 12592 (authorizing the FBI to create an index of DNA records and analyses of persons convicted of crimes, from crime scenes, and from unidentified bodies, and setting forth restricted access and confidentiality protections); Privacy Act of 1974, 5 U.S.C. § 552a; System of Records, 81 Fed. Reg. 27,284, 27,284 (May 5, 2016) (to be codified at 28 C.F.R. § 16) (announcing updates to the FBI's fingerprint identification system records and opening plans for public notice and comment).

226. *See supra* Parts I.A.1–A.3 (discussing data collection and suspect-identification techniques conducted by private entities).

227. *Compare*, *e.g.*, *Quality Assurance Standards for Forensic DNA Testing Laboratories*, FBI (Sept. 1, 2011), https://ucr.fbi.gov/lab/biometric-analysis/codis/quality-assurance-standards-for-forensic-dna-testing-laboratories [https://perma.cc/C5Y3-QEP5] (prescribing quality assurance standards and testing for laboratories conducting DNA forensic testing or using CODIS), *with* Gov't Response to Discovery (*Chatrie*), *supra* note 224, at *4 (explaining the government lacks access to the requested information about Google's Sensorvault data and employees).

identification. A use case is the task that an application is specified to perform.[228] In software engineering, designing for a use case is important to ensure quality control and operational effectiveness in achieving intended purposes.[229] In contrast, misuse cases hijack the intended functionality of a system, with unintended consequences.[230] Designing with misuse cases in mind is a strategy to anticipate security and safety threats and conceptualize sources of systemic malfunction.[231]

For Google, the use case of its SensorVault and connected databases is to facilitate targeted advertising and tailor products and services.[232] The goal of consumer data collection is to elicit more engagement and spending and commercialize the data that users reveal in their interactions with a product.[233] Geofence and keyword warrants drawing on consumer data to identify

228. *See* Brian Dobing & Jeffrey Parsons, *The Role of Use Cases in the UML: A Review and Research Agenda* ("A use case is a description of a sequence of actions constituting a complete task or transaction in an application."), *in* SUCCESSFUL SOFTWARE ENGINEERING 111, 111 (Sal Valenti ed., 2002).

229. *See* Saurabh Tiwari & Atul Gupta, *A Systematic Literature Review of Use Case Specifications Research*, 67 INFO. & SOFTWARE TECH. 128, 128–30 (2015) ("Use case models are typically used as the input in various software development activities, so their ability to clearly document correct, coherent, and an understandable set of functional requirements is critically important for the quality of resulting software product.").

230. Ian Alexander, *Misuse Cases: Use Cases with Hostile Intent*, IEEE SOFTWARE, Jan.-Feb. 2003, at 58, 58–59.

231. *See* Guttorm Sindre, *A Look at Misuse Cases for Safety Concerns* (discussing misuse cases as a mechanism for safety analysis), *in* SITUATIONAL METHOD ENGINEERING: FUNDAMENTALS AND EXPERIENCES 252, 253 (Jolita Ralyté et al. eds., 2007).

232. *See* Declaration of Marlo McGriff at 3, United States v. Chatrie, 590 F. Supp. 3d 901 (E.D. Va. Mar. 11, 2020) (No. 19-cr-00130) [hereinafter Declaration of McGriff] (describing Google's use of SensorVault to store user data for targeted advertising); Jennifer Valentino-DeVries, *Google's Sensorvault Is a Boon for Law Enforcement. This Is How It Works.*, N.Y. TIMES (Apr. 13, 2019), https://www.nytimes.com/2019/04/13/technology/google-sensorvault-location-tracking.html (discussing Google's use of SensorVault for advertising purposes).

233. *See, e.g.*, Kindra Cooper, *How Does Google Market Google? By Putting Customer Data at the Forefront*, CUSTOMER CONTACT WEEK DIGIT. (Oct. 21, 2019), https://www.customercontactweekdigital.com/customer-insights-analytics/articles/google-marketing-data [https://perma.cc/97YN-U4LL] ("The masses of data Google collects from user behavior on its web and mobile properties is used not only to inform its own ad campaigns but is sold to third-party advertisers."); Dokyun Lee et al., *Advertising Content and Consumer Engagement on Social Media: Evidence from Facebook*, 64 MGMT. SCI. 5105, 5107 (2018) (discussing strategies to maximize engagement with ads).

potential perpetrators of a crime is plainly a very different use case—potentially an unanticipated misuse case of the data that Google accumulated. Adverse publicity surrounding Google data collection and law enforcement use is contrary to business interests in maximizing profits and shareholder value.[234] The Google policy change in December 2023 to store data on devices and subvert geofence warrant requests is an example of redesigning system specifications to address potential misuse cases.[235]

The use case for facial recognition technology marketed to businesses seeking to track the risk of shoplifting or persons who may pose a security threat may seem more similar to criminal identification than coopting commercial databases. Yet alerting loss prevention or security officers to be vigilant is different than identifying a particularized person for arrest or conviction.[236] Purveyors of facial recognition products advertise that early recognition of potential shoplifters can prevent and reduce losses through early intervention.[237] Facilitating private self-help to disrupt or repel potential security or retail loss threats is different from post hoc identifying a perpetrator of a crime for arrest and prosecution. Privately-employed loss prevention officers or security guards are not subject to the legal standard of reasonable, articulable suspicion for a temporary on-the-scene *Terry* stop, or probable cause for an arrest.[238]

---

234. Malathi Nayak, *All the Ways Google Is Coming Under Fire over Privacy*, BLOOMBERG NEWS (Feb. 28, 2022), https://www.bloomberg.com/news/articles/2022-02-28/all-the-ways-google-is-coming-under-fire-over-privacy-quicktake (discussing market value plunge over privacy backlash and lawsuits regarding Google data collection and privacy incursions).

235. *See, e.g.*, Davey Alba, *Google Will Stop Providing Law Enforcement Data on Which Users Were Near a Crime*, TIME (Dec. 14, 2023), https://time.com/6539416/google-location-history-data-police [https://perma.cc/6AKN-FBDY] (discussing Google's location data storage changes in responses to privacy concerns about law enforcement use).

236. *See, e.g.*, *Retail Solutions*, FACEFIRST, https://www.facefirst.com/retail [https://perma.cc/6LJW-RY9D] (advertising facial recognition technology that detects "habitual shoplifters" and "persons of interest" and notifies designated personnel when these people enter stores "to proactively prevent shoplifting").

237. *Id.*

238. *See* Terry v. Ohio, 392 U.S. 1, 21 (1968) (requiring reasonable, articulable suspicion of a crime before a government actor may stop a person); United States v. Watson, 423 U.S. 411, 415–16 (1976) (requiring probable cause for arrests by government agents).

Each deviation from a use case results in the risk of unanticipated consequences and errors.[239] Because data is used for unintended purposes and without known baselines against which deviations may be measured, the potential risks of error are hard to quantify and anticipate, and may vary depending on the circumstances of a case, location, and suspect.[240] The costs of error are borne by the person targeted based on the law enforcement cooptation of commercial big data, resulting in suboptimal incentives to avoid error and disparate burdens on suspects.[241]

3. Unknown, Highly Variable, and Racially Disparate Risks of Error

Porcha Woodruff was eight months pregnant and getting her children ready for school when police arrested her in front of her family for a carjacking she did not commit, based on a faulty facial recognition match.[242] Robert Julian-Borchak Williams was arrested in front of his wife and two daughters for a crime he did not commit, allegedly shoplifting five timepieces from an upscale boutique in Detroit.[243]

Alonzo Sawyer, fifty-four, spent nine days in jail after a wrongful arrest based on erroneous facial recognition results

239. *Cf.* Robert K. Merton, *The Unanticipated Consequences of Purposive Social Action,* 1 AM. SOCIO. REV. 894, 899 (1936) (discussing how differences in situations lead to deviations from the usual or expected consequences of an act).

240. *Cf.* Itiel E. Dror, *The Error in "Error Rate": Why Error Rates Are So Needed, Yet So Elusive*, 65 J. FORENSIC SCI. 1034, 1035 (2020) (explaining the difficulties of determining error rates, including the need to have a baseline "ground truth" of known fact against which to measure deviations).

241. *Cf.* Aurélie Ouss & Alexander Peysakhovich, *When Punishment Doesn't Pay: Cold Glow and Decisions to Punish*, 58 J.L. & ECON. 525, 635 (2015) (finding in experiments on the imposition of punishment that "[e]xternalizing costs leads to large increases in punishment levels" because the decisionmaker does not internalize the costs).

242. Complaint & Jury Demand at 2, Woodruff v. Detroit, No. 23-CV-11886 (E.D. Mich. Aug. 3, 2023); Kashmir Hill, *Eight Months Pregnant and Arrested After False Facial Recognition Match*, N.Y. TIMES (Aug. 6, 2023), https://www.nytimes.com/2023/08/06/business/facial-recognition-false-arrest.html.

243. *See* Kashmir Hill, *Wrongfully Accused by an Algorithm*, N.Y. TIMES (Aug. 3, 2020), https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html (reporting how Robert Julian-Borchak Williams was arrested in front of his wife and kids for a crime he did not commit, due to error-prone technology used by law enforcement).

that confused him for a younger, shorter man.[244] Georgia resident Randal Reid was wrongly arrested and jailed for nearly a week for a theft in Louisiana—a state he never visited in his life—based on an erroneous facial recognition match.[245] Nijeer Parks spent ten days in jail and paid $5,000 for his defense after he was wrongly identified by facial recognition technology as a person who shoplifted candy and then tried to hit a police officer with a car.[246]

Michael Oliver, twenty-six, lost his job after he was wrongly arrested for theft based on a faulty facial recognition match to a photo of the perpetrator taken on a cell phone.[247] Ousmane Bah, seventeen, was an honors student at Bronx Latin Academy when he was wrongly arrested for a series of thefts from Apple stores based in part on erroneous facial recognition results.[248]

Woodruff, Williams, Sawyer, Reid, Parks, Oliver, and Bah all were Black.[249] Their lived experiences and the impact on their families show the severe harms of racially disparate error rates with facial recognition technology.[250] Almost all facial

244. Khari Johnson, *Face Recognition Software Led to His Arrest. It Was Dead Wrong*, WIRED (Feb. 28, 2023), https://www.wired.com/story/face-recognition-software-led-to-his-arrest-it-was-dead-wrong.

245. Kashmir Hill & Ryan Mac, '*Thousands of Dollars for Something I Didn't Do*,' N.Y. TIMES (Mar. 31, 2023), https://www.nytimes.com/2023/03/31/technology/facial-recognition-false-arrests.html.

246. Kashmir Hill, *Another Arrest, and Jail Time, Due to a Bad Facial Recognition Match*, N.Y. TIMES (Dec. 29, 2020), https://www.nytimes.com/2020/12/29/technology/facial-recognition-misidentify-jail.html.

247. Elaisha Stokes, *Wrongful Arrest Exposes Racial Bias in Facial Recognition Technology,* CBS NEWS (Nov. 19, 2020), https://www.cbsnews.com/news/detroit-facial-recognition-surveillance-camera-racial-bias-crime [https://perma.cc/WJ2T-FTHK].

248. Bah v. Apple Inc., No. 19-cv-3539, 2021 WL 4084500, at *4–5, *9 (S.D.N.Y. Dec. 28, 2021); Complaint & Jury Demand at 2, 15, Bah v. Apple, Inc., No. 21-cv-10897 (D. Mass. dismissed Sept. 27, 2021) [hereinafter Complaint & Jury Demand (*Bah*)].

249. *See* Hill, *supra* note 242 (noting Woodruff is Black); Hill, *supra* note 246 (noting Parks is Black); Stokes, *supra* note 247 (noting Oliver is Black); Hill, *supra* note 243 (noting Williams is Black); Johnson, *supra* note 244 (noting that Sawyer and Reid are Black); Complaint & Jury Demand (*Bah*), *supra* note 248, at *4 (noting Bah is Black).

250. *See, e.g.*, Thaddeus L. Johnson & Natasha N. Johnson, Opinion, *Police Facial Recognition Technology Can't Tell Black People Apart*, SCI. AM. (May 18, 2023), https://www.scientificamerican.com/article/police-facial-recognition-technology-cant-tell-black-people-apart [https://perma.cc/56PW-8UY5] (summarizing findings that police departments that deploy automated facial

recognition algorithms evaluated over three decades perform differently based on the race and the skin tone of the face, among numerous other factors that impact accuracy.[251]

Machine learning algorithms suffer from the difficulties humans have in recognizing faces of other races with variations depending on the algorithm, the dataset on which the algorithm was trained, and other circumstances.[252] Error rates vary with the thresholds of similarity specified for match identification, the image quality and skin tone of the probe photo, gender, and even the illumination or facial brightness in the probe.[253] Because many facial recognition algorithms were trained on celebrity faces, which skew Caucasian and have strong lighting and makeup, algorithms can have poorer recognition accuracy for under-represented groups, and for darker images with poorer lighting.[254] However, some algorithms are actually more accurate in identifying matches for African-Americans (lower false negatives) and worse at identifying a match for Caucasians (higher

---

recognition technology have increased disparities in arrests of Black persons compared to white persons); Thaddeus L. Johnson et al., *Facial Recognition Systems in Policing and Racial Disparities in Arrests*, GOV'T INFO. Q., Oct. 2022, at 1, 1 (finding that the use of facial recognition technology "contributes to greater racial disparity in arrests . . . [and] [t]his relationship was underpinned by statistically meaningful and positive FRT effects on Black arrest rates and negative effects on White rates").

251. Jacqueline G. Cavazos et al., *Accuracy Comparison Across Face Recognition Algorithms: Where Are We on Measuring Race Bias*, 3 IEEE TRANSACTIONS ON BIOMETRICS BEHAV. & IDENTITY SCI. 101, 101 (2021) ("Nearly all of the face recognition algorithms studied over the past 30 years show some performance differences as a function of the race of the face.").

252. *Id.* at 102–03.

253. *See, e.g.*, Haiyu Wu et al., *Face Recognition Accuracy Across Demographics: Shining a Light into the Problem*, 2023 IEEE/CVF CONF. ON COMPUTER VISION AND PATTERN RECOGNITION WORKSHOPS 1041, 1041–43 (finding that illumination and exposure brightness affects accuracy of facial recognition technology); Joy Buolamwini & Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, 81 PROC. MACH. LEARNING RSCH. 1, 6, 10 (2018) (analyzing the intersection between race, gender, and error rate in machine learning algorithms).

254. Gwangbin Bae et al., *DigiFace-1M: 1 Million Digital Face Images for Face Recognition*, 2023 IEEE/CVF WINTER CONF. ON APPLICATIONS COMPUT. VISION 3515, 3516 ("Face recognition models are generally trained and tested on celebrity faces, many of which are taken with strong lighting and make-up. Celebrity faces also have imbalanced racial distribution . . . 84.5% of the faces in CASIA-WebFace are Caucasian faces . . . leading to poor recognition accuracy for the under-represented racial groups.").

false negatives).[255] The risks of being wrongfully identified—a false positive—are higher, however, for Black, Hispanic, and Asian faces—though error rates have been improving with advances in algorithmic development and training datasets.[256] While the risks of wrongful identification are disparately borne, the potential harms of misidentification can befall persons of any race with error risks varying even based on factors such as hairstyle, accessories, or facial anomalies.[257] For example, Murphy, the grandfather who was wrongly misidentified by facial recognition technology and alleges that he was sexually assaulted while jailed, is white.[258]

While there is growing attention to the racially disparate rates of error with facial recognition technology, the problem of variable and unknown error rates permeates other exemplars of the third generation of big data suspect identifications. For geofence methods of suspect identification using Google's SensorVault and connected databases, a person's location is estimated as a radius with varying margins of error based on manifold sources of potential location information, such as GPS, Bluetooth beacons, and cell towers.[259] The margin of error can be wide enough to capture someone who was entirely outside a geofenced area.[260] Communities and areas vary in access to connectivity and the density and quality of connection points, resulting in the phenomenon of what the literature on

255. *See* Krishnapriya et al., *supra* note 187, at 8–9, 12 (discussing algorithms that more accurately identify matches for African-Americans than Caucasians).

256. Sidney Perkowitz, *The Bias in the Machine: Facial Recognition Technology and Racial Disparities*, MIT CASE STUD. SOC. & ETHICAL RESP. COMPUTING, Winter 2021, at 1, 6–8 (discussing higher error rates among Black, Hispanic, and Asian faces and recent findings that show how these disparate error rates could be remedied).

257. *See, e.g.*, Philipp Terhörst et al., *A Comprehensive Study on Face Recognition Biases Beyond Demographics*, 3 IEEE TRANS. ON TECH. & SOC'Y 16, 20–21 (2022) (discussing the presence of non-racial attributes, such as make-up, weight, and hairstyle, with differential error rates).

258. For a discussion of Murphy's case, see *supra* notes 25–32 and accompanying text. *See also* Murphy v. Essilorluxottica USA Inc., No. 2024-03265 (125th Dist. Ct., Harris Cnty., Tex. Jan. 18, 2023).

259. Declaration of McGriff, *supra* note 232, at 8–9, ¶¶ 23–25.

260. *In re* Search of Info. that Is Stored at Premises Controlled by Google, LLC, 542 F. Supp. 3d 1153, 1158 (D. Kan. 2021); *In re* Search of Info. Stored at Premises Controlled by Google, 481 F. Supp. 3d 730, 745 (N.D. Ill. 2020); Declaration of McGriff, *supra* note 232, at 7, ¶ 25.

technological access disparities terms "digital deserts" or "data poverty."[261] For example, rural, poorer areas have less access to strong signal and connectivity.[262] Because locating suspects was not a planned use of the data, there are no systematic baseline studies to evaluate accuracy rates and how they vary depending on circumstances.

Similarly, the probability that a set of search terms will net uninvolved persons rather than the perpetrator varies greatly with the framing of the keywords, as well as the circumstances of a crime. The keyword warrants in the 2018 Austin pipe bombings investigation—among the few to be unsealed—are illustrative.[263] As the community reeled from the deaths and injuries from mailed pipe bombs detonating on people's porches, FBI investigators sought keyword warrants to compel Google, Microsoft, and Yahoo to reveal users who searched the addresses of the homes where the bombs detonated.[264] More problematically overbroad, investigators also sought users who searched general bomb-related terms such as "cardboard" or "package" and "bomb" or "pipe bomb" or "PVC bomb."[265] Plainly, more idly curious but uninvolved people are likely to search terms like cardboard, package, and bomb, than the address of the targeted homes.

Moreover, the secrecy surrounding keyword warrants severely constrains the ability of defendants and the public to know, much less challenge, potentially problematic investigative

261. Janet Delgado et al., *Bias in Algorithms of AI Systems Developed for COVID-19: A Scoping Review*, 19 BIOETHICAL INQUIRY 407, 416 (2022).

262. *See* Janessa M. Graves, *Disparities in Technology and Broadband Internet Access Across Rurality Implications for Health and Education*, 44 FAM. & CMTY. HEALTH 257, 258–59 (2021) (addressing how often rural areas lack adequate cellular coverage and the barriers to broadband access).

263. *See supra* text accompanying notes 92–94 (noting how the keywords were initially sealed in the Austin pipe bombing case and then selected ones were eventually unsealed).

264. Affidavit re Google Searches, *supra* note 82, at 4, ¶ 6; Affidavit re Microsoft Searches, *supra* note 83, at 4, ¶ 6; Affidavit re Yahoo Searches, *supra* note 83, at 4, ¶ 6.

265. *In re* Search of Info. & Recs. Associated with Google Searches for Various Search Terms that Are Stored at Premises Controlled by Google, No. 18-MJ-00191, at 2, ¶ 2 (W.D. Tex. Mar. 19, 2018) (filed under seal) (on file with the Minnesota Law Review).

tactics.[266] Even with the limited release of just a fraction of the keyword warrants that are filed, we know that requests can vary widely in terms of time frame and breadth of keywords, which in turn varies the risk of netting the innocent. The breadth of the time frame has varied from two weeks before the Diol arson-murders to a month before the Austin pipe bombings.[267] The precision of the keywords specified has varied from the address of a targeted home or person to general words related to bombs.[268] The secrecy also means defendants and the public are in the dark about how many people may be affected by a keyword warrant and the returns on searches. The lack of information means tactics evade effective democratic deliberation and regulation, as well as potential defense challenges over the risk of inaccuracies.[269]

II. THE ADVANTAGES OF EVIDENCE LAW REGULATION OVER FOURTH AMENDMENT FETISHISM

The dominant approach to addressing concerns over government use of technology in criminal investigation is to invoke the Fourth Amendment.[270] New investigative strategies that coopt consumer big data controlled by private companies pose a conundrum for the Fourth Amendment, with its focus on government action, privacy, and longstanding third-party exposure

266. *See, e.g.*, Schladebeck, *supra* note 92 ("To make matters worse, police are currently doing this in secret, which insulates the practice from public debate and regulation.").

267. Motion to Suppress Evidence (*Seymour*), *supra* note 74, at 7, ¶ 21; Affidavit re Google Searches, *supra* note 91, at 24, ¶ 2.

268. Response to Motion to Suppress Evidence (*Seymour*), *supra* note 4, at 2; Affidavit of Scott Kibbey, *In re* Search of Info. & Recs. Associated with Google Searches for Various Search Terms that Are Stored at Premises Controlled by Google, No. 180-MJ-00191 (W.D. Tex. Mar. 19, 2018) (filed under seal) (on file with the Minnesota Law Review).

269. *See, e.g.*, Friedman & Ponomarenko, *supra* note 207, at 1846 (discussing the need for democratic deliberation over rules governing policing); *see also*, David Alan Sklansky, *Police and Democracy*, 103 MICH. L. REV. 1699, 1802 (2005) (discussing the need for greater attention to structures of democratic oversight over policing).

270. *See, e.g.*, Orin S. Kerr*, Searches and Seizures in a Digital World*, 119 HARV. L. REV. 531, 536–40 (2005) (explaining that the "modern Supreme Court has used the text of the Fourth Amendment to craft a comprehensive set of rules regulating law enforcement" and grappling with how these rules extend to computer searches, seizures, and data).

doctrine.[271] To date, commentators and courts have largely tackled the Fourth Amendment thicket of issues.[272] This Article's aim is to go beyond the Fourth Amendment chorus. First, this Part explains the major doctrinal gaps that present problems with the dominant Fourth Amendment focus.[273] This Part then argues for the advantages of centering evidence law and procedures to address the growing welter of concerns over cooptation of privately held data and algorithms beyond their use case with major potential accuracy concerns and harms.[274]

A. MAJOR GAPS IN THE FOURTH AMENDMENT FIXATION ON PRIVACY

Fourth Amendment first principles illuminate the conundrum posed by investigative strategies drawing on the data that we pour onto the Internet or share with private companies. At the threshold, for the Fourth Amendment to apply, there must be a search or seizure by a government actor.[275] Since the landmark Supreme Court decision in *Katz v. United States* in 1967, a search is an intrusion on a reasonable expectation of privacy.[276] According to the third-party exposure doctrine, we do not have an expectation of privacy in information we share with a third party—even just a commercial third party providing vital services such as our bank or telecommunications provider.[277] A

271. *See, e.g.*, Fan, *supra* note 8, at 884, 911 (discussing the struggle to adapt the Fourth Amendment and originalist interpretation of the text to big data searches).

272. *See* sources cited *supra* note 8.

273. *See infra* Part II.A.

274. *See infra* Part II.B.

275. *See* Skinner v. Ry. Lab. Execs.' Ass'n, 489 U.S. 602, 614 (1989) ("As our precedents indicate, not every governmental interference with an individual's freedom of movement raises such constitutional concerns that there is a seizure of the person."); *see also* New Jersey v. T.L.O., 469 U.S. 325, 334–35 (1985) ("But this Court has never limited the Amendment's prohibition on unreasonable searches and seizures to operations conducted by the police. Rather, the Court has long spoken of the Fourth Amendment's strictures as restraints imposed upon 'governmental action'—that is, 'upon the activities of sovereign authority.'" (quoting Burdeau v. McDowell, 256 U.S. 465, 475 (1921))).

276. 389 U.S. 347, 351 (1967); *see also* Smith v. Maryland, 442 U.S. 735, 740 (1979).

277. *See Smith*, 442 U.S. 745 (holding there is no reasonable expectation of privacy to phone numbers recorded in a pen registry); *see also* United States v. Miller, 425 U.S. 435 (1976) (holding the respondent possessed no Fourth Amendment interest in bank records).

second additional standard for a Fourth Amendment search is a physical intrusion on a person's property to gather information—a more limited definition that leaves out the myriad ways police can access our data without physically intruding on our property.[278]

After decades of ostrich-like neglect of the reality that some of the most powerful information about us are shared in the era of smartphones, the Internet, and the Internet of Things, the Supreme Court adjusted the third-party exposure doctrine—somewhat.[279] In *Carpenter v. United States*, the Supreme Court dipped a toe in recognizing modern realities by holding that tracking our movements via cell site location information for seven days or more violates our reasonable expectations of privacy even though we share that data with the cell phone company.[280] The Court did not overrule its prior third-party exposure decisions on bank and phone records.[281] Rather *Carpenter* ruled that because of the "unique nature of cell phone location records, the fact that the information is held by a third party does not by itself overcome the user's claim to Fourth Amendment protection."[282] Access to prolonged retrospective tracking data that could reveal intimate information such as sexual, political, or religious associations without the historical cost barriers of real-time gumshoe work exceeded our reasonable expectations of privacy in the whole of our physical movements.[283] The Court reasoned that cell phone users do not "voluntarily 'assume[] the risk'" of revealing a "comprehensive dossier" of

278. *See* Florida v. Jardines, 569 U.S. 1, 5–6 (2013) ("The Amendment establishes a simple baseline, one that for much of our history formed the exclusive basis for its protections: When 'the Government obtains information by physically intruding' on persons, houses, papers, or effects, 'a "search" within the original meaning of the Fourth Amendment' has 'undoubtedly occurred.'" (quoting United States v. Jones, 565 U.S. 400, 406 n.3 (2012))).

279. *See* Carpenter v. United States, 138 S. Ct. 2206, 2217–19, 2220–21 (2018) (holding that a warrantless acquisition of a person's seven-day cell-site records violated the Fourth Amendment).

280. *See id.* at 2217 n.3 ("It is sufficient for our purposes today to hold that accessing seven days of cell site location information constitutes a Fourth Amendment search.").

281. *See id.* at 2220 ("We do not disturb the application of *Smith* and *Miller* or call into question conventional surveillance techniques and tools, such as security cameras.").

282. *Id.* at 2217.

283. *Id.* at 2218–19.

movements by simply using a cell phone, which must connect to cell phone towers to deliver services.[284]

Lest readers assume a major break with the third-party exposure doctrine, however, *Carpenter* stated, "Our decision today is a narrow one."[285] *Carpenter* then detailed unanswered questions regarding real-time tracking using cell site location information, "tower dumps" of all devices that connected to a cell tower at a particular time interval, and retrospective tracking of less than seven days in duration.[286] *Carpenter* also took pains to emphasize it did not disturb the continued vitality of the third-party exposure doctrine explained in *Smith* and *Miller*; nor the long-settled government practice of using subpoenas for records held by third parties based on mere relevance "in the overwhelming majority of investigations"; nor the use of security cameras; nor government use of other business records, including those that might "incidentally reveal location information."[287]

How about Fourth Amendment seizure regulation? A Fourth Amendment–regulated seizure is a meaningful interference with the possessory interest in property by the government.[288] The data that companies accumulate and compile on us and millions of other consumers for commercial purposes is their property—not ours.[289] Moreover, data is not a depletable asset—it can be used to identify the perpetrator of a murder without meaningfully interfering in the power of Google to use that data to target us with a Hawaii vacation or Olive Garden ads, depending on our location and search predilections.[290] So we can bracket and lay aside Fourth Amendment seizures doctrine, which courts have done, largely focusing on the conundrum of Fourth Amendment search regulation.

Carpenter's landmark—yet cautiously caveated—adjustment to the third-party exposure doctrine was issued in 2018.[291] As of 2024, the Supreme Court has not further revised the third-party exposure doctrine for our networked era. Meanwhile

---

284. *Id.* at 2220 (alteration in original) (quoting Smith v. Maryland, 442 U.S. 735, 745 (1979)).

285. *Id.*

286. *Id.*

287. *Id.* at 2220, 2222.

288. United States v. Jacobsen, 466 U.S. 109, 113 (1984).

289. Fan, *supra* note 17, at 1465–68.

290. *Id.* at 1477–78.

291. *See generally* Carpenter v. United States, 138 S. Ct. 2206 (2018).

surveillance strategies used by law enforcement have advanced so rapidly that cell site location information seems antiquated. Who needs approximate cell site data from cell phone towers when there is Google location history information garnered from much more precise GPS signals, nearby Wi-Fi networks, and Bluetooth devices in addition to cell towers?[292] Cell site location information can be imprecise to the extent of dozens to hundreds of urban city blocks, rather than the margin of error of mere meters for Google location information.[293] In areas with a strong GPS signal, Google location information can be estimated to within twenty meters or less.[294] In rural areas cell site location information is even worse, because of more limited cell site tower coverage—up to forty times less precise than the already-imprecise nature of cell tower location triangulation in urban areas.[295]

Scholars and lower courts are left puzzling over how *Carpenter* applies to new search strategies and digital data.[296] In his empirical study of how lower courts have wrestled to apply "notoriously vague" *Carpenter*, Matthew Tokson aptly observed that "Fourth Amendment law is in flux."[297] Analyzing federal and state court decisions attempting to apply *Carpenter* between June 2018 through March 2021, Tokson found courts in conflict over how narrowly to construe *Carpenter*.[298] Federal courts tended to narrowly construe *Carpenter* and attempt to maintain the status quo predating the decision.[299] About 36.1% of the decisions that Tokson classified as substantively interpreting *Carpenter* were resolved by the court altogether dodging a decision on the open Fourth Amendment question by applying the good

292. *See* Brief of Amicus Curiae (Google), *supra* note 114, at 10 (explaining how Google location information is substantially more precise than the cell site location information in *Carpenter*).

293. *Id.*

294. Declaration of McGriff, *supra* note 232, at 2, ¶ 12.

295. Brief of Amicus Curiae (Google), *supra* note 114, at 10.

296. *See, e.g.*, Tokson, *supra* note 9, at 1791, 1795, 1804 (chronicling how lower courts have wrestled to apply *Carpenter*'s "notoriously vague" standard); Paul Ohm, *supra* note 9, at 394–408 (discussing the many unanswered questions in *Carpenter's* wake).

297. Tokson, *supra* note 9, at 1791.

298. *See id.* at 1811–19 (noting differences in application in federal and state courts and possible factors).

299. *Id.* at 1813–14.

faith exception to the exclusionary rule.[300] The good faith exception permits courts to dispose of a case without even deciding the Fourth Amendment issue because even if there might have been a violation, there is no exclusionary remedy because of police reliance on a warrant, or on then-existing law.[301] Even excluding the large number of cases in the sample resolved on the good faith exception, only 21.2% of the cases found a Fourth Amendment search.[302]

Consider, for example, five appellate cases wrestling with whether and how the Fourth Amendment applies to geofence and keyword warrants.[303] The first federal appellate court to rule on whether geofence data is a Fourth Amendment-regulated search, the Fourth Circuit in *Chatrie v. United States* held that because there is no reasonable expectation of privacy in location data shared with Google, there is no Fourth Amendment-regulated search.[304] *Chatrie* held that the Fourth Amendment's protections do not apply when police ask Google to share user location data for a two-hour interval around the time and place of a bank robbery.[305] The *Chatrie* Court applied the third-party exposure doctrine in which a person lacks a reasonable expectation of privacy in information shared with third parties—even if the information is just shared with the bank, phone company, or Google.[306] Over a dissent, the *Chatrie* majority declined to expand the Supreme Court's decision in *Carpenter* on cell site

300. *See id.* at 1809, 1840 (reporting that 144 out of the 399 decisions substantively applying *Carpenter* resolved the case on the good faith exception "without directly resolving the search issue").

301. *See, e.g.*, United States v. Leon, 468 U.S. 897, 922 (1984) (declining to suppress evidence due to reasonable reliance on a warrant); William J. Mertens & Silas Wasserstrom, *The Good Faith Exception to the Exclusionary Rule: Deregulating the Police and Derailing the Law*, 70 GEO. L.J. 365, 371 (1981) (explaining that the good faith exception diminishes the strength of the Fourth Amendment's protections because there would be no remedy even if there was a violation—and perhaps not even a ruling on whether there was a violation).

302. Tokson, *supra* note 9, at 1812.

303. United States v. Chatrie, 107 F.4th 319, 324, 330–32 (4th Cir. 2024), *reh'g en banc granted*, 2024 WL 4648102 (4th Cir. Nov. 1, 2024); Commonwealth v. Kurtz, 294 A.3d 509, 522–24 (Pa. Super. Ct. 2023), *appeal docketed*, 306 A.3d 1287 (Pa. Oct. 30, 2023); United States v. Smith, 110 F.4th 817, 833–34 (5th Cir. 2024); People v. Seymour, 536 P.3d 1260, 1280 (Colo. 2023).

304. *Chatrie*, 107 F.4th at 324, 330–32.

305. *Id.*

306. *Id.* (applying *United States v. Miller*, 425 U.S. 435, 443 (1976) (bank records), and *Smith v. Maryland*, 442 U.S. 735, 740 (1979) (phone records)).

location data of seven days or more duration to two hours of location data shared by a user with Google. The *Chatrie* Court noted that two hours was far less pervasive than the seven days or more at issue in *Carpenter*—and the user opted to activate location services in *Chatrie*, thereby voluntarily disclosing location information with Google.[307] The continued vitality of this ruling is an open question because a majority of the active Fourth Circuit judges granted en banc review of *Chatrie*, with a decision pending at this writing.[308]

Similarly to the *Chatrie* appellate panel, a Pennsylvania appellate court considering a challenge to a keyword warrant used to identify a serial kidnapper and rapist of five women applied the third-party exposure doctrine to keyword data shared with Google.[309] The keyword warrant sought users that searched for sexual assault survivor K.M.'s address shortly before the crime.[310] The Pennsylvania Superior Court in *Commonwealth v. Kurtz* held that the defendant had no reasonable expectation of privacy in his search queries of K.M.'s residential address before he kidnapped her and raped her because he shared the search queries with Google.[311] Moreover, even assuming arguendo that there was a Fourth Amendment-protected privacy interest, the *Kurtz* Court held that the keyword warrant sufficiently established probable cause.[312] The affidavit established a fair probability that the perpetrator of the kidnapping and sexual assault planned the attack at K.M.'s remote residence after stalking her and likely searched for her address in preparation.[313] The case

307. *Id.* at 337–38 ("Here, we find that Chatrie—unlike Carpenter—did voluntarily expose his Location History to Google.").

308 *See generally* United States v. Chatrie, 590 F. Supp. 3d 901 (E.D. Va. 2022), *aff'd*, 107 F.4th 319 (4th Cir. 2024), *reh'g en banc granted by*, 2024 WL 4648102 (4th Cir. Nov. 1, 2024).

309. *Kurtz*, 294 A.3d at 522–24.

310. *See id.* at 524 ("Trooper Follmer thus sought to compel Google to turn over the IP addresses for users who conducted searches for K.M.'s name or home address in the seven days prior to her attack.").

311. *See id.* at 522 ("By typing in his search query into the search engine and pressing enter, Appellant affirmatively turned over the contents of his search to Google . . . and voluntarily relinquished his privacy interest.").

312. *See id.* at 524 (finding no error in the "issuing authority's probable cause determination").

313. *See id.* at 523–24 (describing the circumstances set forth in the warrant).

currently is pending review by the Pennsylvania Supreme Court.[314]

The Court of Appeals of Minnesota took an intermediate position on the constitutional propriety of geofence warrants in *State v. Contreras-Sanchez*.[315] Holding that geofence warrants that are sufficiently well-drawn based on the circumstances of the crime can satisfy the Fourth Amendment, the Minnesota Court of Appeals rejected an attempt to categorically ban geofence warrants by likening them to historically reviled general warrants.[316] The case involved an attempt to solve the murder of a person dumped in a remote rural field with his hands bound behind his back and a nail shoved into his foot.[317] Though the body was severely decomposed, investigators ultimately determined the slain person was a Minneapolis man reported missing by his family.[318] The investigators obtained a geofence warrant directing Google to disclose any users present in the area where the killer dumped the body for the nearly four-week period between when the family last saw the victim and the discovery of his badly decomposed body.[319] Google produced an anonymized list of twelve devices, with one device clearly more suspicious based on its prolonged stay in the area around the time the body was likely dumped.[320] By a second warrant, investigators obtained the identity details of a single device.[321] The Minnesota Court of Appeals upheld the geofence warrant as sufficiently precisely drawn based on the circumstances of the crime to satisfy the Fourth Amendment's requirements of probable cause and particularization.[322]

A ruling by a Fifth Circuit panel represents the opposite extreme from cases holding that the Fourth Amendment does not apply at all to data we share with Google, and also a contrary

314. Commonwealth v. Kurtz, 306 A.3d 1287 (Pa. 2023) (unpublished table decision).

315. State v. Contreras-Sanchez, 5 N.W.3d 151, 163–64 (Minn. Ct. App. 2024), *review granted* (May 29, 2024).

316. *Id*. at 164–65.

317. *Id*. at 156; State v. Contreras-Sanchez, No. 27-CR-21-20626, 2022 WL 22864500, ¶ 2(a) (D. Ct. Minn., Hennepin Cnty., 4th Jud. Dist. May 2, 2022) (order denying motion for joinder).

318. *Contreras-Sanchez*, 5 N.W.3d at 156.

319. *Id*. at 156–57.

320. *Id*. at 158–59.

321. *Id.*

322. *Id*. at 165–67.

view to the intermediate moderate view of the Minnesota Court of Appeals.[323] About a month after the Fourth Circuit's decision in *Chatrie*, the Fifth Circuit issued a decision disagreeing with *Chatrie*'s reading of the Fourth Amendment and *Carpenter*.[324] In *United States v. Smith*, the Fifth Circuit held that even though a geofence of one or two hours is more temporally limited than the seven days or more of cell site location information addressed in *Carpenter*, the risk of pervasive searches that concerned the Supreme Court in *Carpenter* remains.[325] Moreover, even though a user opts into Location Services to benefit from Google products and services, the *Smith* Court believed this consent was insufficiently informed and voluntary.[326] The Fifth Circuit ruled that obtaining geofence data from Google constitutes a Fourth Amendment search and that a geofence warrant is akin to the historically reviled general warrants and writs of assistance authorizing "general, exploratory rummaging" barred by the Fourth Amendment.[327] Under the good faith exception to the exclusionary rule, however, the evidence obtained pursuant to the geofence warrant was not excludable because the officers relied on the geofence warrant issued by the magistrate judge.[328]

The good faith exception has been an avenue to avoid wading too far into the murky Fourth Amendment morass after *Carpenter*, as illustrated by one of the first major published decisions on keyword warrants.[329] The Colorado Supreme Court in *People v. Seymour* held that the police engaged in a Fourth Amendment–regulated search by issuing keyword warrants to search engine providers Google, Yahoo, and Microsoft, seeking users

323. United States v. Smith, 110 F.4th 817, 833–34 (5th Cir. 2024).

324. *Id.*

325. *Id.* at 835 ("In short, geofence location data is invasive for Fourth Amendment purposes. Of particular concern is the fact that a geofence will retroactively track anyone with Location History enabled, regardless of whether a particular individual is suspicious or moving within an area that is typically granted Fourth Amendment protection.").

326. *Id.* ("Given the ubiquity—and necessity—in the digital age of entrusting corporations like Google, Microsoft, and Apple with highly sensitive information, the notion that users voluntarily relinquish their right to privacy and 'assume the risk' of this information being divulged to law enforcement is dubious.").

327. *Id.* at 836–37 (quoting Coolidge v. New Hampshire, 403 U.S. 443, 467 (1971)).

328. *Id.* at 840.

329. *See generally* People v. Seymour, 536 P.3d 1260 (Colo. 2023).

who googled the address of a home burned during arson-murders shortly before the crime.[330] The difficult question was whether the keyword warrants satisfied the Fourth Amendment's particularity and probable cause requirements. The *Seymour* Court ruled the geofence warrants were sufficiently particularized in specifying the narrow set of keywords and time frame sought and the place to be searched.[331] The *Seymour* Court declined to decide whether the Fourth Amendment required probable cause particularized to a particular user, however, because even if the government lost on that issue, the results of the search would be admissible under the good faith exception to the exclusionary rule.[332]

This quintet of the first appellate courts to address the constitutionality of geofence and keyword warrants shows how the Fourth Amendment after *Carpenter* remains riddled with confusion, gaps, and dodged questions regarding digital searches. Courts disagree on whether the Fourth Amendment applies at all to location data of far less than seven days duration shared with Google.[333] Courts also use the good faith exception to the exclusionary rule to admit evidence despite a search the court found violated the Fourth Amendment—or to avoid the murkiest Fourth Amendment questions.[334]

B. HOW EVIDENCE LAW IS BETTER SUITED TO ADDRESS ACCURACY, RELIABILITY, AND CHANGING TECHNOLOGY

Rather than hoping the fifty-four over-strained and vague words of the Fourth Amendment will address the concerns posed by big data searches, an overlooked body of law is a better theoretical and pragmatic source of protections. Evidence law and procedures are not limited by the Fourth Amendment's myopic focus on privacy, third-party exposure, and government

330. *Id.* at 1273.

331. *Id.* at 1275–77.

332. *Id.* at 1278.

333. *See supra* text accompanying notes 303–328 (noting how the Fourth and Fifth Circuits have taken different positions on how geofencing and location tracking implicates the Fourth Amendment).

334. *See* United States v. Smith, 110 F.4th 817, 840 (5th Cir. 2024); *Seymour*, 536 P.3d at 1278.

action.[335] Rather, evidence law is theoretically and pragmatically oriented toward addressing accuracy, reliability, and access to information used as evidence.[336] Evidence law also is capacious and elastic enough to accommodate policy concerns such as privacy, but is not moored to just that predominant lens.[337] Moreover, courts operating as evidentiary gatekeepers may conduct rigorous fact-finding about the science and technology behind evidence, such as big data suspect identifications.[338]

The "driving theory" of evidence law is ensuring accuracy in factfinding.[339] To serve that goal, trial judges serve a central screening role and apply rules of evidentiary preclusion to ensure the reliability of evidence to guide fact-finders toward accurate decisions.[340] The strong discretion accorded trial judges as gatekeepers of evidence is a subconstitutional way to address the risk of problematic evidentiary inputs that are not screened out by constitutional criminal procedure, which is focused on governmental wrongdoing from a privacy lens.[341] Scholars have sought to further expand strong screening and scrutiny power by judges beyond scientific evidence to other controversial error-prone

335. *See supra* text accompanying notes 285–287, 291–295 (overviewing the limitations of relying on the Fourth Amendment to provide adequate protection over data concerns).

336. *See, e.g.*, Michael S. Pardo, *supra* note 13, at 559 (discussing how accuracy and allocation of the risks of error are central concerns of evidence law and procedures); Darryl K. Brown, *The Decline of Defense Counsel and the Rise of Accuracy in Criminal Adjudication*, 93 CALIF. L. REV. 1585, 1622 (2005) (discussing the import of pretrial discovery in evidentiary proceedings to improving reliability).

337. *See, e.g.*, Jeffries v. Nix, 912 F.2d 982, 986 (8th Cir. 1990) (noting that a purpose of the evidentiary rules protecting sexual assault victims is protecting privacy); Michelle J. Anderson, *From Chastity Requirement to Sexuality License: Sexual Consent and a New Rape Shield Law*, 70 GEO. WASH. L. REV. 51, 88–93 (2002) (explaining how evidence rules adapted to promulgate rape shield laws exist to protect victim privacy, among other policy concerns).

338. *See, e.g.*, State v. Henderson, 27 A.3d 872, 877–78 (N.J. 2011) (detailing how in a case challenging the reliability of eyewitness identification, the court appointed a special master to evaluate the scientific evidence on eyewitness reliability, resulting in review of hundreds of scientific studies, testimony by seven experts, and yielding more than 2,000 transcript pages).

339. Allen, *supra* note 13, at 632.

340. *Id.*

341. *See* Jennifer E. Laurin, *Quasi-Inquisitorialism: Accounting for Deference in Pretrial Criminal Procedure*, 90 NOTRE DAME L. REV. 783, 799 (2014) (discussing "corrective mechanisms outside criminal procedure doctrine to both better equip trial processes to sort out bad evidentiary inputs").

forms of evidence such as eyewitness and informant testimony.[342]

Evidence rules and procedures also have a public-facing component, in reassuring the public about the accuracy and acceptability of the judgment.[343] Thus evidence law is aimed not just at maximizing the probability of accuracy, but also the appearance of justice in a publicly palatable sense.[344] Some evidence rules, such as the prohibition against convicting based just on high statistical probability of guilt, reflect this sensitivity to moral considerations in evidence.[345] As evidenced by modern changes, such as evidentiary protections for victims of sexual assault, evidence law also is capacious enough to accommodate public policy concerns and evolving norms in addition to the overarching aims of facilitating accuracy in fact-finding and evidentiary reliability.[346]

Evidence law also is more responsive and nimbler than cumbersome constitutional law in adapting to technological change and public opinion because of how evidence rules are generated and interpreted. An Advisory Committee on Evidence Rules appointed by the Chief Justice of the U.S. Supreme Court has the power to propose amendments to the Federal Rules of Evidence in light of experience in the courts and new challenges.[347] The advisory committee engages in "a continuous study of the

342. *See* Sandra Guerra Thompson, *Judicial Gatekeeping of Police-Generated Witness Testimony*, 102 J. CRIM. L. & CRIMINOLOGY 329, 339 (2012) ("Among the proposals put forth by academics and advocacy groups is the proposal that trial courts expand their judicial gatekeeping role to include pretrial reliability reviews of police-generated witness testimony.").

343. *See* Charles Nesson, *The Evidence or the Event? On Judicial Proof and the Acceptability of Verdicts*, 98 HARV. L. REV. 1357, 1369 (1985) ("Toward this end, the value of promoting acceptable verdicts has considerable explanatory power, illuminating and ac- counting for many existing evidentiary rules and procedures.").

344. *Cf.* Clayton Littlejohn, *Truth, Knowledge, and the Standard of Proof in Criminal Law*, 197 SYNTHESE 5253, 5262–63 (2020) (making epistemic arguments against using only statistical evidence for conviction).

345. *See id.* at 5263 ("If the reader's moral sensibilities are anything like mine, they'll be troubled by the suggestion that we can punish using statistical evidence.").

346. *See, e.g.*, Anderson, *supra* note 337, at 88–93 (describing the evolution of evidentiary protections for victims of sexual assault).

347. *See, e.g.*, Memorandum from Patrick J. Schiltz, Chair, Advisory Comm. on Evidence Rules, to John D. Bates, Chair, Standing Comm. on Rules of Practice & Proc. (May 15, 2022) [hereinafter Schiltz Memorandum] (reporting final approval to three amendments to the Federal Rules of Evidence).

operation and effect" of the rules and considers "suggestions and recommendations received from any source, new statutes and court decisions affecting the rules, and legal commentary."[348] The proposed amendments are promulgated for public notice and comment by judges, practitioners, and the public generally.[349] The Judicial Conference's Committee on Rules of Practice and Procedure, also referred to as the Standing Committee, reviews the public comments and proposed amendments and, if it agrees, sends the amendments to the U.S. Supreme Court.[350] If the Court concurs, it can order the revisions, which become effective unless Congress enacts legislation that rejects, changes, or defers the proposed amendments.[351] States also tend to utilize advisory committees on the rules of evidence to recommend updates, often influenced by the Federal Rules of Evidence.[352]

The continuous updating of federal and state rules of evidence by expert committees attuned to practice in the field is a sharp contrast to the severely constrained certiorari process of the U.S. Supreme Court and glacial updating of Fourth Amendment doctrine for technological change. The Court grants certiorari on about one percent of the petitions for review it receives.[353] Over the decades, even as digital technologies proliferated, the probability of the Supreme Court granting certiorari plummeted.[354] The Supreme Court has called for judicial restraint rather than impose constitutional straitjackets on fast-evolving

348. *Procedures for the Judicial Conference's Committee on Rules of Practice and Procedure and Its Advisory Rules Committees*, U.S. CTS. § 440.20.10 (May 27, 2022), https://www.uscourts.gov/sites/default/files/guide-vol01-ch04-sec440_procedures_for_rules_cmtes_1.pdf [https://perma.cc/XUA3-E6EW].

349. 28 U.S.C. § 2071(b); *see also, e.g.*, Schiltz Memorandum, *supra* note 347, at 2–9 (reporting out amendments for public notice and comment).

350. *How the Rulemaking Process Works*, U.S. CTS., https://www.uscourts.gov/rules-policies/about-rulemaking-process/how-rulemaking-process-works [https://perma.cc/9AJB-QXPH].

351. 28 U.S.C. § 2072; *How the Rulemaking Process Works*, *supra* note 350.

352. *See Rules: Federal Rules of Evidence*, FED. JUD. CTR., https://www.fjc.gov/history/work-courts/rules-federal-rules-evidence [https://perma.cc/Q5EN-V347] (noting adoptions by states); *see, e.g.*, MINN. R. EVID. 703 (noting review and incorporation of Federal Rules of Evidence).

353. Barry P. McDonald, *SCOTUS's Shadiest Shadow Docket*, 56 WAKE FOREST L. REV. 1021, 1040 (2021).

354. *Id.*

societal and technological change.[355] Courts interpreting spare constitutional text are ill-suited to conduct the factual inquiry needed to frame nuanced rules for the complexities of fast-changing technology.[356]

Moreover, when the legislature intervenes in framing evidentiary procedures, the social concerns embedded in pretrial screening mechanisms can democratically reflect local morality and norms—for better or worse depending on whether one shares those views. For example, reflecting a strong solicitude for self-defense claimants, the Florida legislature required pretrial screening of prosecutions involving self-defense claims, requiring the prosecution to prove by clear and convincing evidence that the defendant was not acting in self-defense.[357] To take another example, the rise of evidentiary rules for the protection of sexual assault claimants reflected the changing consciousness of communities across the nation about protecting the privacy of survivors and redressing the harms of attacking survivors.[358] In contrast, Fourth Amendment jurisprudence, influenced heavily by the norms and assumptions of the Justices of the U.S. Supreme Court, is oft-critiqued as out of step with public understandings and norms.[359]

---

355. *See, e.g.*, City of Ontario v. Quon, 560 U.S. 746, 759 (2010) ("The judiciary risks error by elaborating too fully on the Fourth Amendment implications of emerging technology before its role in society has become clear.").

356. *See, e.g.*, Indus. Indem. Co. v. Alaska, 669 P.2d 561, 563 (Alaska 1983) ("[C]ourts must not intrude into realms of policy exceeding their institutional competence."); *cf.* Allison Orr Larsen, *Constitutional Law in an Age of Alternative Facts*, 93 N.Y.U. L. REV. 175, 181 (2018) (noting "[t]he traditional claim is that courts do not have the same factfinding tools as legislatures"—though challenging the conventional wisdom).

357. *See* Love v. State, 286 So. 3d 177, 180 (Fla. 2019) (discussing the self-defense pre-trial screening legislation). For a critique, see, for example, Eric Ruben, *Self-Defense Exceptionalism and the Immunization of Private Violence*, 96 S. CAL. L. REV. 509, 534 (2023).

358. Anderson, *supra* note 337, at 88–93; *see* Marilyn J. Ireland, *Reform Rape Legislation: A New Standard of Sexual Responsibility*, 49 U. COLO. L. REV. 185, 192 (1978) ("Legislative reform of evidentiary provisions protects the victim from unnecessary intrusions into his or her privacy.").

359. *See, e.g.*, Christopher Slobogin & Joseph E. Schumacher, *Reasonable Expectations of Privacy and Autonomy in Fourth Amendment Cases: An Empirical Look at "Understandings Recognized and Permitted by Society*," 42 DUKE L.J. 727, 732 (1993) (offering empirical evidence of the gap between the understandings and assumptions of the U.S. Supreme Court Justices and that of the public).

Evidentiary procedures also offer pretrial screening mechanisms to evaluate reliability and potentially force information disclosure as the price of evidentiary admissibility.[360] While pretrial screenings, like any pragmatic flexible process relying on strong judicial discretion, are not a systemic cure-all, in the criminal context, it provides some accountability and check on police and prosecutorial power.[361] Pretrial evidentiary proceedings also have adapted to address computerized or digital evidence, algorithms, and databases, and the need for information to evaluate reliability.[362] For example, Judge Valerie Caproni in the Southern District of New York ordered the disclosure of the source code of the Forensic Statistical Tool for probabilistic genotyping under protective order in pretrial *Daubert* hearings on the reliability of expert DNA evidence.[363] After journalists at *ProPublica* filed a motion arguing that there was public interest in revealing the source code, Judge Caproni lifted the protective order, permitting the source code to be posted online for the defense and public in future cases.[364] The outcome illustrates how evidentiary procedures and trial judges situated closer to technological and scientific debates can accomplish far more than the current Fourth Amendment stalemate in addressing concerns over secrecy, reliability, and the need to test privately-controlled data and algorithmic products.

---

360. *See, e.g.*, John P. Manard Jr. et al., *Case Strategy and Trial Management*, *in* TOXIC TORT LITIGATION 167, 200 (Arthur F. Foerster & Christine Gregorski Rolph eds., 2d ed. 2013) ("Because public disclosure of discovery material is subject to the discretion of the trial court under Rule 26(c), confidential materials filed in connection with pretrial discovery remain protected so long as 'good cause' has been shown.").

361. *See* Jessica A. Roth, *Informant Witnesses and the Risk of Wrongful Convictions*, 53 AM. CRIM. L. REV. 737, 745 (2016) (noting that while not perfect, pretrial screening in myriad criminal evidentiary contexts is valuable, "lending some accountability and feedback to police and prosecutors").

362. *See, e.g.*, Hannah Bloch-Wehba, *Access to Algorithms*, 88 FORDHAM L. REV. 1265, 1314 (2020) (discussing cases); MANUAL FOR COMPLEX LITIGATION § 21.446 (3d ed. 1995) ("Issues concerning accuracy and reliability of computerized evidence, including necessary discovery, should be addressed during pretrial proceedings and not raised for the first time at trial.").

363. Protective Order Regarding the Confidentiality of the Forensic Statistical Tool Source Code & Related Documents, United States v. Johnson, No. 15-cr-00565 (S.D.N.Y. Jul. 18, 2016).

364. Jason Tashea, *Code of Science Defense Lawyers Want to Peek Behind the Curtain of Probabilistic Genotyping*, ABA J., Dec. 2017, at 18, 19.

III. UPDATING EVIDENCE RULES FOR BIG DATA-BASED SUSPECT IDENTIFICATIONS

This section presents three proposals for how evidence rules and procedures can address the accuracy concerns, secrecy, and risk of harms surrounding big data-based suspect identifications.[365] The first protection is updating the conceptions of probable cause for big data suspect identifications to include limits against using geofence or keyword warrant returns or a facial recognition match as probable cause for an arrest absent corroboration.[366] The second set of safeguards includes rigorous pretrial notice, disclosure and reliability inquiries for big data suspect identification techniques before the results may be referenced as evidence.[367] Third, this Article proposes offering expert witnesses on the error risks of big data suspect identifications if the fruits of such tactics are presented at trial.[368] The advantages of evidentiary rule-making and implementation is that these ideas can be implemented through multiple avenues and actors: by recommendation of advisory committees, by legislation, or by the exercise of the broad discretion of trial judges as gatekeepers and screeners of evidence.[369]

A. UPDATING PROBABLE CAUSE FOR ARRESTS BASED ON BIG DATA SUSPECT IDENTIFICATIONS

One of the gravest harms of big data suspect identifications is wrongful arrests based on faulty results.[370] To protect against the risks of wrongful arrests and the tunnel vision that arises from a seemingly data-based identification, the concept of probable cause must be updated such that the results of a geofence or keyword warrant or a facial recognition hit cannot alone constitute probable cause for an arrest. There must be further corroboration. If the results cannot constitute probable cause absent corroboration, *a fortiori*, the evidence, even if admissible at trial, cannot alone constitute proof beyond a reasonable doubt to support a criminal conviction, a much higher standard than

365. For a discussion of the concerns and potential harms, see *supra* Parts I.B.1–B.3.

366. *See infra* Part III.A.

367. *See infra* Part III.B.

368. *See infra* Part III.C.

369. See *supra* text accompanying notes 339–356 on evidentiary rulemaking and the gatekeeping role of judges.

370. *See* discussion *supra* Part I.B.3.

probable cause.[371] Such an update incentivizes further law enforcement investigation, disrupting the tunnel vision—the fixation on one target while ignoring contrary information—exemplified in the case of Jorge Molina, where police ignored evidence that his stepfather rather than Molina was the murderer.[372]

The proposal is practicable, as illustrated by on-the-ground policy changes emerging in some jurisdictions. For example, responding to extensive public debate and controversy, the New York Police Department's policy on facial recognition software now explains that facial recognition match "shall be treated as an investigative lead only" and cannot constitute probable cause for an arrest or search warrant; rather "[c]orroborating information must be developed through additional investigation by the assigned investigator."[373]

Though a start, police department policies alone are insufficient because wrongful arrests can still arise if law enforcement officers arrest a person notwithstanding departmental policy.[374] Showing that a wrongful arrest was in violation of police departmental policy alone is insufficient to make out a civil rights claim that survives summary judgment based on qualified immunity for officers, which requires a violation of clearly established rights.[375]

---

371. *See, e.g.*, Fisher v. Jordan, 91 F.4th 419, 428 (6th Cir. 2024) ("[P]robable cause demands even less than that necessary to establish a *prima facie* case at trial, let alone to convict." (internal citation omitted)); Commonwealth v. Taylor, 69 Pa. D. & C.2d 748, 752 (Pa. Ct. Com. Pl. 1974) ("[F]light standing alone is not sufficient to form probable cause for an arrest, and therefore, *a fortiori*, is insufficient to prove guilt beyond a reasonable doubt." (internal citation omitted)).

372. *See supra* notes 33–42 and accompanying text; *see also, e.g.*, Brian Reichart, *Tunnel Vision: Causes, Effects, and Mitigation Strategies*, 45 HOFSTRA L. REV. 451, 451 (2017) (explaining the role of tunnel vision on wrongful convictions).

373. *Patrol Guide: Facial Recognition Technology*, N.Y.C. POLICE DEP'T, (Mar. 12, 2020), https://www.nyc.gov/assets/nypd/downloads/pdf/nypd-facial-recognition-patrol-guide.pdf [https://perma.cc/JB9B-AR5E].

374. *See, e.g.*, Perryman v. City of Bloomington, 704 F. Supp. 3d 961, 967 (D. Minn. 2023) (discussing wrongful arrest based on facial recognition match in violation of municipal policy requiring independent verification).

375. *See, e.g.*, Davis v. Scherer, 468 U.S. 183, 196 (1984) (showing a violation of regulations alone is insufficient to overcome qualified immunity); Backlund v. Barnhart, 778 F.2d 1386, 1390 n.5 (9th Cir. 1985) (determining that a violation of internal departmental policy is insufficient to overcome qualified immunity).

Moreover, a prohibition in a police department manual may insulate municipalities because it means plaintiffs cannot meet the demanding *Monell* standard for lawsuits against municipalities.[376] In *Monell v. New York City Department of Social Services*, the Supreme Court ruled that cities are "persons" for purposes of civil rights lawsuits under 42 U.S.C. § 1983.[377] To state a claim, however, civil rights plaintiffs must show that a violation arose because of the execution of a municipality's policy or custom, and not just from an employee's malfeasance.[378] The Supreme Court has extended the demanding *Monell* requirement to lawsuits seeking injunctive or declarative relief, as well as suits for damages.[379] The lack of remedies for erroneous data-based suspect identifications means that the serious harms suffered by people like Harvey Eugene Murphy, Jr. or Jorge Molina may go unredressed and undeterred, risking harm to more wrongly identified persons.[380]

The wrongful arrest of Kylese Perryman in violation of police department policy illustrates the need for explicit codification by evidentiary rule or via judicial interpretation.[381] Perryman was wrongly arrested as the perpetrator of a series of violent robberies based on a misidentification using facial recognition technology deployed by Hennepin County detectives.[382] Yet the police training manual for Hennepin County required independent verification of a facial recognition match, and provided that a software hit is a lead generator, not a sufficient basis for a positive identification of a suspect.[383] The U.S. District of Minnesota dismissed Perryman's civil rights claims against Hennepin County brought under 42 U.S.C. § 1983, explaining he failed to state a sufficient *Monell* claim for municipal liability.[384] The fact that the arresting detectives acted in violation of the

376. *See, e.g.*, Perryman v. City of Bloomington, 704 F. Supp. 3d 961, 967 (D. Minn. 2023) (discussing how a wrongful arrest based on facial recognition match was in violation of a municipal policy requiring independent verification).

377. 436 U.S. 658, 690 (1978); 42 U.S.C. § 1983.

378. *Monell*, 436 U.S. at 690–91; *see also* Los Angeles County v. Humphries, 562 U.S. 29, 36 (2010).

379. *Monell,* 436 U.S. at 690.

380. See the discussion of the cases of mistaken data-based suspect identifications, *supra* text accompanying notes 1–2.

381. *Perryman*, 704 F. Supp. 3d at 966–67.

382. *Id.*

383. *Id.*

384. *Id.* at 973.

police department policy actually insulated the municipality because it prevented Perryman from meeting the *Monell* standard that the violation arose from official policy or custom.[385] This outcome illustrates the necessity of explicit codification or judicial pronouncement that arrests cannot be predicated on a big data suspect identification alone to deter wrongful arrests and offer a basis for sanctions, including civil rights lawsuits.

The Seventh Circuit's decision on ShotSpotter alerts in *United States v. Rickmon* offers an example of how judges can update standards of proof to address the use of controversial technology.[386] Deployed by police departments in more than 100 U.S. cities, ShotSpotter uses a network of acoustic sensors to detect the sound of gunfire so that police can rapidly respond in hopes of apprehending likely perpetrators and helping the wounded.[387] The technology is controversial because the sensors are concentrated in communities suffering from firearms violence, which disproportionately are disadvantaged BIPOC communities.[388] The concentration of acoustic surveillance rouses concern that ShotSpotter alerts heighten racial disproportionality in persons who police stop and arrest.[389]

Rickmon involved the arrest of Terrill Rickmon after a ShotSpotter alert regarding two gunshots at 4:40 a.m. in Peoria, Illinois.[390] The detective responding to the ShotSpotter alert saw only one vehicle leaving the scene and stopped the vehicle.[391] He obtained consent to search the vehicle and found a nine-millimeter handgun under the passenger seat occupied by defendant Rickmon, who had a prior felony conviction.[392] Charged with being a felon in possession of a firearm, Rickmon argued that the police stop that led to the discovery of the firearm was unlawful because a ShotSpotter alert is insufficient to constitute

385. *Id.* at 966.

386. 952 F.3d 876, 881 (7th Cir. 2020) (evaluating the totality of the circumstances).

387. Mitchell L. Doucette et al., *Impact of ShotSpotter Technology on Firearm Homicides and Arrests Among Large Metropolitan Counties: A Longitudinal Analysis, 1999-2016*, 98 J. URB. HEALTH 609, 610 (2021).

388. Christopher Slobogin & Sarah Brayne, *Surveillance Technologies and Constitutional Law*, 6 ANN. REV. CRIMINOLOGY 219, 224 (2022).

389. *Id.*

390. *Rickmon*, 952 F.3d at 879.

391. *Id.*

392. *Id.*

reasonable articulable suspicion for a *Terry* stop.[393] Police must meet the standard of reasonable articulable suspicion, a standard lower than probable cause, to engage in a brief, on-the-scene detention called a *Terry* stop, which is short of a full-blown arrest.[394] The Seventh Circuit agreed with Rickmon's contention "that ShotSpotter, standing on its own, should not allow police officers to stop a vehicle in the immediate vicinity of a gunfire report without any individualized suspicion of that vehicle."[395] The Seventh Circuit further questioned "whether a single ShotSpotter alert would amount to reasonable suspicion," likening such an alert to an anonymous tipster in terms of low quantum of proof.[396] Ultimately the Seventh Circuit upheld the stop on the facts of the case because the ShotSpotter alerts were corroborated by multiple 911 calls reporting shots fired and fleeing suspects and the car was the only vehicle on the sole street leading away from the scene.[397]

In addition to requiring corroboration, rules should delineate what does not suffice as corroboration. Importantly, the corroboration cannot be tainted by, or largely the product of, the initial big data suspect identification technique. Two cases illustrate the problem. First in *People v. Reyes*, police investigating package thefts from a mail room used an image of the thief captured on security cameras as a probe for a facial recognition search.[398] The facial recognition software returned a single hit—for the defendant Reyes, whose mugshot was in the database.[399] Departmental policy forbade police from using a facial recognition match as probable cause for an arrest.[400] So the detective viewed the surveillance footage that generated the probe for the facial recognition match, and Reyes's mug shot and tattoos and attested that he recognized Reyes as the person depicted in the surveillance footage.[401] The detectives who arrested Reyes also viewed the surveillance photo montage and attested they

393. *Id.* at 881.
394. Terry v. Ohio, 392 U.S. 1, 21 (1968).
395. *Rickmon*, 952 F.3d at 881.
396. *Id.* at 881–82.
397. *Id.* at 883–84.
398. 133 N.Y.S.3d 433, 434–35 (N.Y. Sup. Ct. 2020).
399. *Id.* at 434.
400. *See Patrol Guide: Facial Recognition Technology*, *supra* note 373.
401. *Reyes*, 133 N.Y.S.3d at 434–35.

recognized Reyes from the photo montage.[402] This farcical end-run around departmental policy that a facial recognition match alone cannot constitute probable cause should not have sufficed because police were acting with tunnel vision focus on Reyes prompted by the match.[403] The match led detectives to believe Reyes was the perpetrator and any identification would be tainted from this confirmation bias.[404]

The nightmare that befell Harvey Eugene Murphy, Jr., also illustrates the harm that can ensue from tainted corroboration.[405] Recall the harrowing allegations in Murphy's civil rights suit at the outset of the article.[406] He was wrongly arrested for robbery based on a faulty facial recognition hit combined with an eyewitness identification tainted by the prior algorithmic identification.[407] The leading cause of wrongful convictions, eyewitness identifications are notoriously malleable yet powerfully persuasive to a jury.[408] Traumatic events, such as being held at gunpoint and focusing on a weapon, can render eyewitness identifications particularly error-prone.[409] Cross-racial identifications where the eyewitness is identifying a person of another race also are fraught with error.[410] Subsequent research has found that low-confidence eyewitness identifications are particularly risky

402. *Id.*

403. For a discussion of the problem of tunnel vision and wrongful convictions, see, for example, Findley & Scott, *supra* note 39, at 292–94.

404. *Cf.* D. Kim Rossmo & Jocelyn M. Pollock, *Confirmation Bias and Other Systemic Causes of Wrongful Convictions: A Sentinel Events Perspective*, 11 NE. U. L. REV. 790, 810–14 (2019) (discussing the role of confirmation bias in heightening the risk of faulty eyewitness identifications and wrongful convictions).

405. *See supra* notes 25–32 and accompanying text.

406. *See supra* notes 25–32 and accompanying text.

407. *See supra* notes 25–32 and accompanying text.

408. *See, e.g.*, Perry v. New Hampshire, 565 U.S. 228, 245 (2012) (acknowledging studies revealing that faulty eyewitness identifications are a leading cause of wrongful convictions); United States v. Wade, 388 U.S. 218, 228 (1967) ("The identification of strangers is proverbially untrustworthy. The hazards of such testimony are established by a formidable number of instances in the records of English and American trials."); Richard A. Wise, et al., *An Examination of the Causes and Solutions to Eyewitness Error*, FRONTIERS PSYCH., Aug. 2014, at 1, 1 (discussing studies on eyewitness suggestibility and the malleability of memory).

409. Jonathan M. Fawcett et al., *Of Guns and Geese: A Meta-Analytic Review of the 'Weapon Focus' Literature*, 19 PSYCH. CRIME & L. 35, 36–38, 54–59 (2011).

410. John Paul Wilson et al., *The Cross-Race Effect and Eyewitness Identification: How to Improve Recognition and Reduce Decision Errors in Eyewitness Situations*, 7 SOC. ISSUES & POL'Y REV. 83, 87–89 (2013).

and the driving cause of wrongful convictions.[411] Confidence levels also are prone to suggestion and taint—which can arise if the eyewitness identification is influenced by knowledge of a facial recognition hit.[412] Just as two wrongs do not make a right, an eyewitness identification tainted by a facial recognition match should not suffice as corroboration for an arrest.

B. PRETRIAL NOTICE AND DISCLOSURE ENABLING CHALLENGES TO BIG DATA IDENTIFICATIONS

The secrecy surrounding the use of big data-based search techniques is another major challenge to ensuring reliability and a fair opportunity to mount a defense.[413] There are at least two main forms of secrecy. The first potential secrecy concern is nondisclosure that a big data search technique led the police to focus on the defendant—or even constituted a basis of an arrest.[414] A second type of secrecy concern is nondisclosure of sufficient information about the technology to permit effective defense challenge to admission of the suspect identification results.[415] This section proposes robust notice, disclosure, and pretrial evidentiary screening for reliability to address both types of secrecy concerns surrounding controversial big data suspect identification strategies.

---

411. John T. Wixted, *Time to Exonerate Eyewitness Memory*, FORENSIC SCI. INT'L, Nov. 2018, at 1, 2.

412. *See* Rachel Leigh Greenspan & Elizabeth F. Loftus, *Eyewitness Confidence Malleability: Misinformation as Post-Identification Feedback*, 44 LAW & HUM. BEHAV. 194, 195 (2020) ("[C]onfidence is malleable and a variety of factors from the time of the crime until the conclusion of a case can influence a witness' confidence in their identification.").

413. *See supra* notes 92–104, 266–69 and accompanying text.

414. *See, e.g.*, Amster & Diehl, *supra* note 8, at 2508–09 (discussing how police attempt to prevent Google from alerting users about law enforcement access to data); Natalie Ram, *Innovating Criminal Justice*, 112 NW. U. L. REV. 659, 666–68 (2018) (describing controversies over failure to disclose use of Stingray cell site simulators in police investigations).

415. *See, e.g.*, State v. Pickett, 246 A.3d 279, 287 (N.J. Super. Ct. App. Div. 2021) (discussing argument that without the source code of probabilistic genotyping software, the defense could not effectively challenge the reliability of the process or conclusions); Ram, *supra* note 414, at 690–91 (discussing the problems with nondisclosure about algorithms and source code).

To challenge a technology that led to a client's arrest, defense attorneys must first know the technology was used.[416] Yet when it comes to suspect identifications using privately-controlled data and products, contractual nondisclosure provisions, trade secrets law, and police nondisclosure requests pose multiple potential barriers to defense discovery.[417] For example, defense attorneys face hurdles in even realizing that facial recognition technology was used as a lead to identify their client.[418] The charging documents might only mention an eyewitness identification—and not the use of facial recognition technology, which might have led to law enforcement tunnel vision and even a taint of the eyewitness identification.[419] The problem of trying to detect whether facial recognition technology was used against a client is acute enough that defense attorneys even try to educate colleagues about signs that they need to dig into the question despite nondisclosure.[420]

Enterprising reporters using Freedom of Information Act requests and state public disclosure laws unearthed another infamous example of contractual secrecy surrounding law enforcement use of cell site simulator devices, also often referred to as a "Stingray."[421] The device simulates a cell tower and forcibly connects with cell phones in the area so that law enforcement can search unique cell phone identifiers for a target phone.[422] The manufacturer of the device, Harris Corporation, insisted on secrecy, and even attempted to block a Freedom of Information

416. Amster & Diehl, *supra* note 8, at 2508–09; C. Justin Brown & Kasha M Leese, *Stingray Devices Usher in a New Fourth Amendment Battleground*, CHAMPION, June 2015, at 12, 13.

417. Ram, *supra* note 414, at 666–68.

418. Kaitlin Jackson, *Challenging Facial Recognition Software in Criminal Court*, CHAMPION, July 2019, at 14, 16.

419. *Id.* See also *supra* notes 403–12 and accompanying text on the problems of tunnel vision and tainted eyewitness identifications that are not independent of a facial recognition match.

420. Jackson, *supra* note 418, at 16–17.

421. Larry Greenemeier, *What is the Big Secret Surrounding Stingray Surveillance*, SCI. AM. (June 25, 2015), https://www.scientificamerican.com/article/what-is-the-big-secret-surrounding-stingray-surveillance [https://perma.cc/PN5D-HQ6S].

422. Ellen Nakashima, *FBI Clarifies Rules on Secretive Cellphone-Tracking Devices*, WASH. POST (May 14, 2015), https://www.washingtonpost.com/world/national-security/fbi-clarifies-rules-on-secretive-cellphone-tracking-devices/2015/05/14/655b4696-f914-11e4-a13c-193b1241d51a_story.html.

request for device manuals.[423] The FBI required that state and local police departments maintain secrecy in agreements where they share the devices.[424] Many defense attorneys, and even the presiding judge, never realized that law enforcement used a Stingray in their case—effectively preventing challenges to a technological tool used to construct a criminal case against their client.[425]

Where law enforcement agents access privately held data, such as Google's motherlode of location or search data, the request can include a nondisclosure request that Google refrain from alerting users about the data access.[426] Sample language from a geofence warrant in a murder investigation includes an order to Google to refrain from notifying users that their data was released to law enforcement because of "the sensitivity of this on-going criminal investigation," and the risk that notification "could compromise this investigation as well as the safety of law enforcement officers participating in the investigation."[427]

Nondisclosure time frames in reported examples have varied, with examples as long as ninety days and six months.[428] Orders sealing the use of keyword warrants also present severe discovery and challenge problems.[429] It is hard to even know how many such warrants have issued, much less how many users

---

423. Elizabeth E. Joh, *The Undue Influence of Surveillance Technology Companies on Policing*, 92 N.Y.U. L. REV. ONLINE 19, 22 (2017).

424. Nakashima, *supra* note 422.

425. *See, e.g.*, Spencer McCandless, Note, *Stingray Confidential*, 85 GEO. WASH. L. REV. 993, 996 (2017); Brad Heath, *Police Secretly Track Cellphones to Solve Routine Crimes*, USA TODAY (Aug. 24, 2015), http://www.usatoday.com/story/news/2015/08/23/baltimore-police-stingray-cell-surveillance/31994181 [https://perma.cc/R5PR-ZZAY].

426. Amster & Diehl, *supra* note 8, at 2508–09.

427. Search Warrant, County/Circuit Court of the Fifteenth Judicial Circuit for Palm Beach County, Florida (May 9, 2018) (on file with the Minnesota Law Review).

428. Amanda Lamb, *Scene of a Crime? Raleigh Police Searched Google Accounts as Part of Downtown Fire Probe*, WRAL NEWS (July 13, 2018), https://www.wral.com/scene-of-a-crime-raleigh-polic-esearch-google-accounts-as-part-of-downtown-fire-probe/17340984 [https://perma.cc/YHP2-JWXT] (ninety days); Tony Webster, *How Did the Police Know You Were Near a Crime Scene? Google Told Them*, MINN. PUB. RADIO NEWS (Feb. 7, 2019), https://www.mprnews.org/story/2019/02/07/google-location-police-search-warrants [https://perma.cc/ZWG7-E5GN] (six months).

429. *See supra* notes 92, 104, 266–69 and accompanying text (discussing the secrecy of big data-based search techniques).

were affected and or the successful rate of return on such warrants.

Police contracts with private technology vendors also may contain nondisclosure agreements regarding the algorithms or source code used because companies require secrecy to maintain trade secret protection for their products.[430] An example of secrecy surrounding facial recognition technology arises from Bronx Defenders attorneys who kept the client's identity secret because of a plea deal.[431] The case arose when a man stole a package of socks and fled after flashing a box cutter at a loss prevention officer attempting to stop him.[432] Police investigators uploaded still images of the sock lifter to facial recognition software and obtained a match.[433] Investigators then texted the photo of the match to the loss prevention officer who served as the sole eyewitness with the note, "Is this the guy?"[434] After that highly suggestive procedure, the loss prevention officer texted back identifying the match photo as the perpetrator, thereby identifying the defendant.[435]

Because the facial recognition match underlay the case, including the eyewitness identification, defense attorneys filed discovery requests seeking information about the facial recognition technology.[436] The government insisted that the arrest was based on eyewitness identification, not the facial recognition technology and fought disclosure of any details about the technology.[437] Prosecutors argued that the NYPD was merely a user and not the owner nor creator of the software and any disclosure would violate trade secret protection for the software.[438] Ultimately, the dispute was swept under the rug and any disclosure avoided with an irresistible plea deal reducing felony charges to a misdemeanor.[439]

---

430. Deborah Won, Note, *The Missing Algorithm: Safeguarding* Brady *Against the Rise of Trade Secrecy in Policing*, 120 MICH. L. REV. 157, 173–74, 190 (2021); Ram, *supra* note 414, at 666–68.

431. Mike Hayes, *'Is This the Guy?,'* APPEAL (Aug. 20, 2019), https://theappeal.org/is-this-the-guy [https://perma.cc/ZWG7-E5GN].

432. *Id.*

433. *Id.*

434. *Id.*

435. *Id.*

436. *Id.*

437. *Id.*

438. *Id.*

439. *Id.*

Defendants have the right to a "meaningful opportunity to present a complete defense" rooted in the Fourteenth Amendment Due Process Clause and the Sixth Amendment.[440] When witnesses against the defendant are human, defendants have a right to impeachment material relevant to testing the reliability of that witness.[441] Concerned about the risks of eyewitness misidentification, some jurisdictions also are creating rights of notice, disclosure, and opportunity to challenge identifications of the defendant.[442]

Yet when a key part of building the case against a defendant was generated by big data analytics, defendants are often denied the right to know the nature of that technological evidence, much less any impeachment material about risks of inaccuracy.[443] Moreover, even jurisdictions like New York that have been progressive about creating rights to notice and opportunities to challenge identification procedures have construed that right to exclude technologically-based identifications.[444] Recall the case of Luis Reyes, where NYPD detectives engaged in an end-run around the internal policy forbidding basing arrests on facial recognition hits by looking at his mugshot obtained via a facial recognition hit and comparing it with security video footage.[445] Reyes never had a chance to challenge the reliability of his identification under article 710 of New York's Criminal Procedure Law, which provides for notice and a pretrial opportunity to challenge identification procedures for being unduly suggestive.[446] The trial court held there was no "identification procedure"

440. Holmes v. South Carolina, 547 U.S. 319, 324 (2006) (quoting Crane v. Kentucky, 476 U.S. 683, 690 (1986)).

441. *See* Giglio v. United States, 405 U.S. 150, 153 (1972) (discussing how allowing evidence to be introduced uncorrected is "incompatible with 'rudimentary demands of justice'" (quoting Mooney v. Holohan, 294 U.S. 103, 112 (1935))).

442. *E.g.*, N.Y. CRIM. PROC. LAW §§ 710.20, 710.30 (McKinney 2024) (creating rights to notice and opportunity to challenge evidence that a defendant was observed at the scene of the crime or identified via "pictorial, photographic, electronic, filmed or video recorded reproduction"); New Jersey v. Henderson, 27 A.3d 872, 919 (N.J. 2011) (holding that a defendant is entitled to a hearing to test variables pertaining to eyewitness reliability if the defendant produces some evidence that a "system variable" produced by criminal justice processing was suggestive).

443. *See supra* notes 416–35 and accompanying text.

444. People v. Reyes, 133 N.Y.S.3d 433, 434–35 (N.Y. Sup. Ct. 2020).

445. *See supra* notes 398–404 and accompanying text.

446. N.Y. CRIM. PRO. LAW § 710.20, 710.30 (McKinney 2024).

within the meaning of the law because the detective comparing the mug shot to the security video was not akin to the eyewitness identifications contemplated by the law.[447]

What happened to Reyes shows the need to explicitly extend pretrial notice, disclosure, and reliability hearings to technologically-assisted suspect identifications that are the fruit of big data analytics.

The power of trial judges to order disclosure regarding suspect identification technologies to facilitate the right to mount a defense is exemplified by the appellate decision of the Superior Court of New Jersey in *New Jersey v. Arteaga*.[448] *Arteaga* involved the use of facial recognition technology to identify a robbery suspect and facilitate eyewitness identifications.[449] Investigators first sought facial recognition analysis from the New Jersey Regional Operations Intelligence Center.[450] A New Jersey analyst advised that he could not find a match but he could try again if provided better-quality images.[451] The robbery investigators instead turned to the New York Police Department's Facial Identification Section, which used their technology to identify the defendant, Francisco Arteaga.[452] Investigators showed two eyewitnesses a six-pack photo array of Arteaga and five other fillers.[453] Both eyewitnesses picked Arteaga as the likely perpetrator.[454]

Indicted on robbery, assault, and weapons charges, Arteaga's defense attorney sought discovery of the testing and operation of the facial recognition technology used to identify him.[455] He sought the name and manufacturer of the software, the source code for the facial recognition algorithms, testing results and error rates, the surveillance image used as a probe to query the software, the confidence scores for possible matches, other potential matches returned, and the qualifications of the analyst who ran the software, among other facts.[456] Prosecutors provided

447. *Reyes*, 133 N.Y.S.3d at 435.
448. New Jersey v. Arteaga, 296 A.3d 542, 554 (N.J. Super. 2023).
449. *Id*. at 545–46.
450. *Id*.
451. *Id*.
452. *Id*.
453. *Id*. at 546.
454. *Id*.
455. *Id*. at 546–47.
456. *Id*.

some of the requested materials, including the match report, the still images used as probes, the first ten possible matches for each probe (with confidence scores), and the analyst's notes.[457] The motions judge denied the defense discovery request for other data the prosecution refused to disclose.[458]

On appeal of the discovery denial, the Superior Court reversed and ordered discovery, noting the motion judge may use protective orders and in camera review for any proprietary or otherwise confidential material.[459] The *Arteaga* court framed its ruling in terms of the prosecution's obligation to release potentially exculpatory evidence, noting that "the items sought by the defense have a direct link to testing FRT's reliability and bear on defendant's guilt or innocence."[460] *Arteaga* ruled that the defendant's Due Process rights would be violated by denial of access to the requested information about the technology, which constituted "raw materials integral to the building of an effective defense."[461]

To delve into the complexities of the technology, courts evaluating evidentiary reliability can use Special Masters or consider expert testimony and an array of scientific and technological studies. The searching reliability inquiries conducted by the New Jersey and Oregon Supreme Courts in the context of evaluating eyewitness testimony reliability offer examples.[462] Concerns about disclosure of trade secrets can be addressed via protective order. Pioneering courts have ordered disclosure of source codes of proprietary technology such as the TrueAllele probabilistic genotyping software under protective order and the sky did not fall, nor the business lose its lucrative product.[463] To address

457. *Id*. at 549.

458. *Id*. at 550–51.

459. *Id*. at 557–58.

460. *Id*. at 558.

461. *Id*. at 554 (quoting Ake v. Okla., 470 U.S. 68, 77 (1985)).

462. State v. Henderson, 27 A.3d 872, 877–78 (N.J. 2011) (detailing how in a case challenging the reliability of eyewitness identification, the court appointed a special master to evaluate the scientific evidence on eyewitness reliability, resulting in review of hundreds of scientific studies, testimony by seven experts, and yielding more than 2,000 transcript pages); State v. Lawson, 291 P.3d 673, 696 (Or. 2012) (en banc) (conducting extensive review of the scientific literature and data on eyewitness reliability).

463. *See, e.g.,* State v. Pickett, 246 A.3d 279, 299–300 (Sup. Ct. N.J. App. Div. 2021) (ordering disclosure of source code of Cybergenetics' TrueAllele under protective order and noting other similar decisions); Protective Order Regarding

potential police end-runs such as that in Reyes's case, the notice, disclosure, and reliability rules should apply whenever the government aims to use any evidentiary fruits traceable to a big data suspect identification.[464]

C. EXPERT WITNESSES TO EDUCATE JURIES ABOUT THE FALLIBILITY OF BIG DATA IDENTIFICATIONS

Big data suspect identifications are part of the rise of proof generated by machines and machine learning that have the aura of scientific and technological conclusiveness, yet present the risk of opaque untested errors.[465] The products of machine learning and big data searches carry the appearance of objectivity before fact-finders.[466] The very inscrutability and seemingly novel nature of big data methods add to the mystique of infallibility, resulting in the risk that the evidence will be overweighted or insufficiently scrutinized.[467] To properly contextualize any big data suspect identifications referenced at trial, this part argues for a best practice of introducing expert witness testimony to contextualize the risk of error.

Expert witness testimony is potentially admissible when "the expert's scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue."[468] Expert witness testimony about factors

---

the Confidentiality of the Forensic Statistical Tool Source Code & Related Documents, United States v. Johnson, No. 15-cr-00565 (S.D.N.Y. July 18, 2016) (ordering disclosure of the software source code for the Forensic Statistical Tool for probabilistic genotyping).

464. *See supra* notes 398–404 and accompanying text.

465. *See, e.g.*, Andrea Roth, *Machine Testimony*, 126 YALE L.J. 1972, 1976–77, 1984–90 (2017) (noting that while the "shift from human- to machine-generated proof has, on the whole, enhanced accuracy and objectivity in fact finding," machine-generated evidence present inscrutable "black box" algorithmic process and evade safeguards subject to hearsay by humans).

466. *See, e.g.*, State v. Brown, 337 N.W.2d 507, 511 (Iowa 1983) (noting "the mystique which may attach to 'objective' machines and quantitative evidence").

467. *See, e.g.*, Elizabeth L. Crooke & Brian D. Depew, *Expert Judgment*, L.A. LAW., Dec. 2012, at 24, 28 ("The reason for judicial caution for novel devices and processes is the misleading aura of infallibility potentially created by a machine."); Ned Miltenberg, *Out of the Fire and into the Fryeing Pan or Back to the Future*, TRIAL, Mar. 2001, at 18, 23 (cautioning regarding "supposedly perfect (and perfectly definitive) mechanical 'black boxes'—that is, machines, techniques, and devices").

468. *E.g.*, FED. R. EVID. 702(a); DEL. R. EVID. 702(a); IND. R. EVID. 702(a); PA. R. EVID. 702(b); TEX. R. EVID. 702.

affecting the reliability of evidence not generally known to juries meets the helpfulness requirement.[469] Courts also recognize that expert witness testimony to counteract potential juror misperceptions meets the helpfulness requirement.[470]

A prime example is the growing number of courts admitting expert witnesses to educate juries about the risks of eyewitness misidentification, breaking with the past practice of excluding such expert testimony.[471] As the Oregon Supreme Court in *State v. Lawson* explained, because the variables that heighten the risk of eyewitness misidentification "are either unknown to the average juror or contrary to common assumptions, expert testimony is one method by which the parties can educate the trier of fact concerning variables that can affect the reliability of eyewitness identification."[472] Courts have admitted expert testimony about factors such as the heightened risk of error in cross-racial identifications, or how "weapon focus" can undermine the accuracy of identifications.[473] Appellate courts have even reversed convictions where trial judges refused to admit expert evidence on cross-racial identifications where the key or sole evidence against a defendant was a cross-racial eyewitness.[474] The

469. State v. Lawson, 291 P.3d 673, 696 (Or. 2012) (en banc).

470. *See, e.g.*, State v. Robinson, 431 N.W.2d 165, 172 (Wis. 1988) (admissibility of expert testimony to correct "misconception that all sexual assault victims are emotional" after an assault); State v. Jensen, 147 N.W.2d 913, 918 (Wis. 1988) ("In this case, [an expert witness's] testimony was relevant because it provided information about behavioral characteristics of child sexual abuse victims that may have been outside the jurors' common experience."); State v. London, 784 N.W.2d 182 (Wis. Ct. App. 2010) (unpublished table decision) (affirming admission of expert witness testimony on how "acting out" can be a sign of childhood sexual abuse).

471. *E.g.*, United States v. Rodriguez-Felix, 450 F.3d 1117, 1124 (10th Cir. 2006).

472. *Lawson*, 291 P.3d at 696; *see also, e.g.,* Commonwealth v. Robinson, 278 A.3d 336, 341 (Pa. Super. 2022) (finding the same).

473. *E.g.*, United States v. Green, 664 F. App'x 193, 196 (3d Cir. 2016) (expert witness on cross-racial identification and "weapon focus"); Workman v. State, 771 S.E.2d 636, 638 (S.C. 2015) (expert witness on "cross-racial identification, memory development during short and traumatic events, and how memories can be influenced when there is a suggestive succeeding event, such as a news broadcast featuring pictures of the defendants").

474. *E.g.*, People v. Boone, 91 N.E.3d 1194, 1199–1200, 1211 (N.Y. 2017) (reversing on many grounds, including cross-racial identification); *see also* State v. Jaime, 233 P.3d 554, 560 (Wash. 2010) (Sanders, J. concurring) (voting to reverse the conviction on the ground the trial court erred in excluding expert

opportunity to cross-examine an eyewitness does not obviate the need for expert witness testimony to counteract the risk that a jury "may overestimate the veracity and reliability of eyewitness identification."[475]

Expert witness testimony similarly is necessary to educate juries about the risks of error and to counteract the risk of juries overweighing big data suspect identifications. Currently, police and prosecutors tend to use the results of geofence, keyword or facial recognition searches to generate leads to other evidence, rather than seek to admit such results in court.[476] Using the techniques as lead generators strategically shields the technology from notice and disclosure requirements.[477] The results of big data analytical techniques are so new and untested that they may not even be admissible in jurisdictions that still follow the rigorous standard of *Frye v. United States* requiring that results be predicated on "well-recognized scientific principle[s] or discovery . . . . sufficiently established to have gained general acceptance in the particular field in which it belongs."[478]

The majority of jurisdictions, however, follow the more permissive standard of *Daubert v. Merrell Dow Pharmaceuticals* jettisoning the general acceptance standard and requiring that judges perform a screening role to ensure that scientific testimony and evidence are reliable as well as relevant.[479] In performing this judicial screening role, trial judges are well situated to ensure adequate notice and disclosure as well as contextualize any results with expert witness testimony to educate juries about reliability. Moreover, if evidence, such as eyewitness testimony, is the fruit of a big data analytical technique, trial judges should admit expert testimony educating jurors about the reliability risks and problems of tunnel vision even if the government

---

testimony on eyewitness identifications where the defendant was "of Hispanic descent" and the eyewitnesses were not Hispanic).

475. *Robinson*, 278 A.3d at 341.

476. *See, e.g.*, People v. Reyes, 133 N.Y.S.3d 433, 436–37 (N.Y. Sup. Ct. 2020) ("To the best of this judge's knowledge, a facial recognition 'match' has never been admitted at a New York criminal trial as evidence that an unknown person in one photo is the known person in another.").

477. *See supra* notes 435–63 and accompanying text.

478. 293 F. 1013, 1014 (D.C. Cir. 1923) (*Frye* standard), *superseded by Federal Rules of Evidence as recognized by* Daubert v. Merrell Dow Pharms., Inc., 509 U.S. 579 (1993); *see also Reyes*, 133 N.Y.S.3d at 436–37 (opining that facial recognition technology match results would fail the *Frye* test).

479. *Daubert*, 509 U.S. at 588–89.

does not directly seek to admit evidence of the big data analytical matches.[480]

CONCLUSION

A new generation of suspect identification techniques drawing on big data analytics presents great power and peril, offering the tantalizing possibility of cracking cold cases such as unsolved murders, but also posing the harms of wrongful arrests.[481] The three big data suspect identification strategies analyzed in this Article, keyword, geofence, and facial recognition technology searches, exemplify this new generation of technology and big data-based suspect identifications.[482] Attempting to address the risks and harms posed by these largely unregulated techniques, courts and commentators have largely focused on further extending the hyper-stretched and vague fifty-four words of the Fourth Amendment regarding government searches and seizures.[483] Yet the Fourth Amendment, with its predominant focus on government invasions of privacy and its third-party exposure doctrine, is an incomplete fit for regulating new suspect identification techniques drawing on privately-controlled big databanks and search algorithms.[484]

This Article advances beyond the chorus parsing and debating the Fourth Amendment and its doctrinal stalemate to center evidence law and procedures as a better-suited body of law for regulating big data suspect identifications.[485] Evidence law protections and procedures focused on reliability, notice, and disclosure can fill the major gaps left unfilled by the fixation on the Fourth Amendment.[486] The Article proposes updating and adapting evidence law safeguards for big data-based suspect identifications and three clusters of reforms.[487] First is expressly providing that probable cause for arrests and proof beyond a reasonable doubt cannot be predicated solely on the results of big data-based suspect identifications.[488] Corroboration is required

480. *See supra* notes 403, 418, 463 and accompanying text.
481. *See supra* notes 25–42, 70–76, 242–51 and accompanying text.
482. *See supra* Parts I.A.1–A.3.
483. *See supra* Part II.A.
484. *See supra* Part II.A.
485. *See supra* Parts II.B., III.
486. *See supra* Part II.B.
487. *See supra* Part III.
488. *See supra* Part III.A.

and evidentiary fruits of a big data analytical technique, such as an eyewitness identification tainted by knowledge of a facial recognition match, does not suffice.[489] Second is rigorous pretrial notice and disclosure requirements to address the secrecy surrounding big data suspect identifications.[490] Third are expert witnesses to educate juries about the reliability concerns and risk of tunnel vision surrounding the evidentiary fruits of big data-based suspect identifications.[491] The goal of these reforms is to preserve the power and potential to crack cold cases with big data suspect identifications while reducing the risks of harms such as wrongful arrests.[492]

489. *See supra* Part III.A.
490. *See supra* Part III.B.
491. *See supra* Part III.C.
492. *See supra* notes 25–42, 70–76, 242–51 and accompanying text.