No. 18-5578

UNITED STATES COURT OF APPEALS
FOR THE SIXTH CIRCUIT

UNITED STATES OF AMERICA

Plaintiff-Appellee,

vs.

WILLIAM J. MILLER,

Defendant-Appellant.

On Appeal from the United States District Court
for the Eastern District of Kentucky
Case No. 2:16-cr-00047-1
The Hon. David L. Bunning

BRIEF OF *AMICUS CURIAE* ELECTRONIC PRIVACY INFORMATION CENTER (EPIC) IN SUPPORT OF APPELLANT

Marc Rotenberg
Counsel of Record
Alan Butler
Electronic Privacy Information Center
1718 Connecticut Avenue, N.W.
Suite 200
Washington, DC 20009
(202) 483-1140

October 17, 2018

UNITED STATES COURT OF APPEALS
FOR THE SIXTH CIRCUIT

Disclosure of Corporate Affiliations and Financial Interest

Sixth Circuit Case Number: 18-5578 Case Name: United States v. Miller

Name of counsel: Alan Butler

Pursuant to 6th Cir. R. 26.1, Electronic Privacy Information Center
Name of Party

makes the following disclosure:

1. Is said party a subsidiary or affiliate of a publicly owned corporation? If Yes, list below the identity of the parent corporation or affiliate and the relationship between it and the named party:

No.

2. Is there a publicly owned corporation, not a party to the appeal, that has a financial interest in the outcome? If yes, list the identity of such corporation and the nature of the financial interest:

No.

CERTIFICATE OF SERVICE

I certify that on October 17, 2018 the foregoing document was served on all parties or their counsel of record through the CM/ECF system if they are registered users or, if they are not, by placing a true and correct copy in the United States mail, postage prepaid, to their address of record.

s/ Alan Butler
1718 Connecticut Ave. NW, Suite 200
Washington, DC 20009

This statement is filed twice: when the appeal is initially opened and later, in the principal briefs, immediately preceding the table of contents. See 6th Cir. R. 26.1 on page 2 of this form.

TABLE OF CONTENTS

TABLE OF AUTHORITIES

CASES

STATUTES

OTHER AUTHORITIES

INTEREST OF AMICUS

The Electronic Privacy Information Center ("EPIC") is a public interest research center in Washington, D.C., established in 1994 to focus public attention on emerging civil liberties issues and to protect privacy, the First Amendment, and other constitutional values.[1]

EPIC routinely participates as *amicus curiae* before the United States Supreme Court and other courts in cases concerning emerging privacy issues, new technologies, and constitutional interests. EPIC has authored several briefs specifically concerning Fourth Amendment standards for searches using new technologies. *See*, *e.g.*, Brief of *Amici Curiae* EPIC et. al, *Carpenter v. United States*, 138 S. Ct. 2206 (2018) (arguing that the Fourth Amendment protects the right against warrantless seizure and search of location data); Brief of *Amici Curiae* EPIC et. al, *Riley v. California*, 134 S. Ct. 2473 (2014) (arguing that warrantless search of a cell phone incident to an arrest is impermissible); Brief of *Amicus Curiae* EPIC, *Florida v. Harris*, 133 S. Ct. 1050 (2013) (arguing that the Government bears the burden of establishing the reliability of techniques used in criminal investigations).

[1] In accordance with Fed. R. App. P. 29, the undersigned states that no monetary contributions were made for the preparation or submission of this brief, and this brief was not authored, in whole or in part, by counsel for a party.

SUMMARY OF ARGUMENT

This case involves the use of a digital forensic technique used to identify images that may be unlawful to possess. This technique analyzes an image and generates a numerical "hash" value that can then be compared to a database of other hash values representing images that have been flagged as containing apparent child pornography. In collaboration with law enforcement agencies, Google uses this technique to automatically scan the private files of internet users. As a consequence, the private files of Gmail users are routinely subject to inspection and analysis, yet neither Google nor the federal agency has revealed the specific nature of the underlying algorithm. Neither Google nor the Government has established the accuracy, reliability, and validity of this technique. Such transparency is necessary because the consequences of an error are severe—automatic referral of a user's data, files, and identity to the National Center for Missing and Exploited Children ("NCMEC") and a subsequent investigation and referral to local law enforcement.

Moreover, the use of this technique for other purposes, e.g. to determine if files contain religious viewpoints, political opinions, or "banned books," would raise profound First Amendment concerns. Indeed, Google is currently facing criticism concerning Project DragonFly, a search engine designed for the Chinese government that enables the identification of materials that China would consider

"politically sensitive." Brian Fung, *Google Really is Trying to Build a Censored Chinese Search Engine, Its CEO Confirms*, Wash. Post (Oct. 16, 2018).[2]

ARGUMENT

As the Supreme Court recently recognized "seismic shifts in digital technology" require a reexamination of existing Fourth Amendment standards. *Carpenter v. United Sates,* 138 S. Ct. 2206, 2219–20 (2018). Regarding law enforcement collection of cell site location information, the Court declined to extend the "third party doctrine," finding instead that the Fourth Amendment required a warrant for this investigative technique. The Court recognized individuals have both "a reasonable expectation of privacy in the whole of their physical movements," and that "law enforcement agents and others would not—and indeed, in the main, simply could not—secretly monitor and catalogue every single movement of an individual's car for a very long period." *Id.* at 2217. Therefore a Fourth Amendment rule permitting suspicionless tracking of suspects in the physical world where such tracking "for any extended period of time was difficult and costly and therefore rarely undertaken," *id.*, could not justify the vast capabilities of digital surveillance.

The same can be said about the private search doctrine as applied to the continuous scanning of private files, stored on computer servers across the country.

[2] https://www.washingtonpost.com/technology/2018/10/16/google-really-is-trying-build-censored-chinese-search-engine-its-ceo-confirms/.

Even if a court determines that the Government's participation in a search was purely passive because it was not involved, the traditional doctrine requires that any "additional invasions of respondents' privacy by the government agent must be tested by the degree to which they exceeded the scope of the private search." *United States v. Jacobsen*, 466 U.S. 109, 131 (1984). In *Jacobsen* the Court held that the Government's warrantless inspection and testing of the contents of a package that had been previously searched by FedEx was permissible because "there was a virtual certainty" that the law enforcement officer's search would not reveal "anything more than he had already been told." *Id*. at 119.

A search is not reasonable under the private search doctrine if (1) it relies on a private company's proprietary technique, (2) the technique is used routinely to search millions of files, and (3) the government does not establish the reliability of the technique in matching images with a virtual certainty. Under the traditional private search doctrine, the Government would clearly be prohibited from opening and inspecting files that had not been previously searched by Google. *United States v. Ackerman*, 831 F.3d 1292, 1294-1304 (10th Cir. 2016) (Gorsuch, J.). But the question in this case is whether the Government has provided sufficient evidence to establish that it is "virtually certain" that the files sent in a CyberTipline Report to the NCMEC were the same as those uploaded by the user.

I. The automated scanning of private files on Google servers affects more than a billion users, which means that any errors in the detection system could have a massive impact on user privacy.

As part of the coordinated effort among electronic communications service providers, the NCMEC, and government investigators, Google scans billions of files to identify suspected contraband. 18 U.S.C.§ 2258A(a)(2). But the sheer volume of data being subjected to these searches, including private files uploaded to cloud storage on the largest platforms, means that the risk of error in the identification or algorithmic matching of these images is significant. *See* Declaration of Cathy McGoff ¶ 4, ECF No. 33-1, *United States v. Miller*, No. 2:16-cv-47, 2017 WL 2705963 (E.D. Ky. Jun. 23, 2017) (describing how Google compares "content uploaded to [their] services" to the hashes of previously flagged images). If a non-contraband image is added to one of these lists by mistake, or if a provider's algorithm falsely matches a non-contraband image with one of the records from its list, many innocent users could immediately have their confidential files relayed to law enforcement and be subject to an intrusive investigation as a result. Strong safeguards are needed to protect the interests of millions of users, especially because an error or mismatch would not likely be subject to judicial review.

Recent studies confirm that e-mail is "the most pervasive form of communication." Radicati Group, Inc., *Email Statistics Report, 2018-2022:*

Executive Summary, at *2 (2018).[3] There are an estimated 3.8 billion e-mail users worldwide in 2018 and the number of accounts is growing at an even faster rate than the number of users. *Id*. at 3.[4] The largest email provider in the world is Google, with more than 1.4 billion Gmail users. Jake Swearingen, *Gmail Gets a Major Face-lift and Productivity Boost, Starting Today*, NY Mag Intelligencer (Apr. 25, 2018).[5] And e-mail only represents a small portion of Google's services. The company controls a wide range of internet services that enable users to upload images and other files. This includes Google Photos, Google Drive, Google Docs, and YouTube. *See* Google, *Our Products* (2018).[6] Google's file storage service, alone, has an estimated 1 billion users worldwide as of 2018. Shoshana Wodinsky, *Google Drive is About to Hit 1 Billion Users*, The Verge (Jul. 25, 2018).[7] This means that Google is scanning millions and millions of images each day.

Given the substantial volume of private files that are subject to Google's image scanning algorithm each day, the potential impact of an error or mismatch is

---

[3] https://www.radicati.com/wp/wp-content/uploads/2018/01/Email_Statistics_Report,_2018-2022_Executive_Summary.pdf.

[4] The recent survey estimates an average 1.75 accounts per user, which is expected to grow steadily over the next four years. *Id*.

[5] http://nymag.com/intelligencer/2018/04/how-to-turn-on-google-gmail-redesign-and-new-features.html.

[6] https://www.google.com/about/products/. Even the Google Search platform relies on user-uploaded images for "reverse image search." Google, *Search for Images with Reverse Image Search* (2018), https://support.google.com/websearch/answer/1325808?hl=en&ref_topic=3180360.

[7] https://www.theverge.com/2018/7/25/17613442/google-drive-one-billion-users.

quite significant and should be treated accordingly. The history of this case and other similar cases reveals the typical process that follows a positive match by Google's scanning algorithm. After an image is flagged, Google automatically submits a CyberTipline Report to NCMEC, which includes:

- the date and time of the incident;
- the e-mail address associated with the user account that uploaded the file;
- the IP address associated with the upload;
- a list of IP addresses used to access the user account (which can go as far back as the original account registration date);
- the filename
- the "categorization" of the image based on an existing rubric; and
- a copy of the image file(s).

CyberTipline Report 5778397 at 1–3, ECF No 33-2, *United States v. Miller*, No. 2:16-cv-47, 2017 WL 2705963 (E.D. Ky. Jun. 23, 2017). The NCMEC system automatically adds information to the report by identifying the following information associated with the user's IP address(es): Country, Region, City, Metro Code, Postal Code, Area Code, Latitude/Longitude, and Internet Service Provider or Organization. *See, e.g.*, *id*. at 4–5. Then the NCMEC staff collect additional information, including "data gathered from searches on publicly-

available, open-source websites" using the account and user identifying information provided in the CyberTipline Report. *See, e.g.*, *id.* at 6. This information gathered by NCMEC can include social media profiles, websites, addresses, and other personal data. *See, e.g.*, *id.* at 6–11. All of this personal data would be collected and then sent to a detective near the user before any person has actually reviewed the image(s) to confirm that they are contraband.[8]

There are at least three types of errors that could trigger the search and production of a Google user's personal data to law enforcement where no contraband image was ever uploaded.

First, a Google employee could mistakenly flag a non-contraband image (record entry error). Depending on a service provider's method for flagging images, it is possible that an employee could either flag the wrong image or mistakenly identify an image as apparent contraband when in fact the image does not contain contraband.

Second, a service provider might erroneously flag an image based on a list of hash values that it received from some other entity (downstream error). The potential for downstream error was previously identified by the court in *United*

[8] *See* McGoff Declaration, *supra*, ¶ 7 ("When Google's product abuse detection system encounters a hash that matches a hash of a known child sexual abuse image, in some cases Google automatically reports the user to NCMEC without re-reviewing the image.").

States v. Keith, 980 F. Supp. 2d 33 (D. Mass. 2013), in a case where AOL's staff had not actually reviewed the original image that was the basis for the hash value.

Third, the flag from the provider's hashing algorithm could be a "false positive" due to the specific image-matching method used (match error). A false positive could, for example, be caused by similarities in the images even if one image contains contraband and the other does not. The likelihood of a mismatch error depends entirely on the specific hashing method used and its false positive rate. For example, certain *file hashing* algorithms are designed "to confirm that when a copy of data is made, the original is unaltered and the copy is identical, bit-for-bit." Richard P. Salgado, *Fourth Amendment Search and the Power of the Hash*, 119 Harv. L. Rev. 38, 38 (2005). But there is no evidence on the record that Google's proprietary *image matching* algorithm matches files bit-for-bit. Other image matching algorithms, including Microsoft PhotoDNA, which is used by NCMEC, identify "similar" images based on analysis of the image's content. Microsoft, *Photo DNA: Step-by-step*.[9]

The Google declaration submitted in this case does not provide sufficient detail to evaluate the reliability and validity of their process for flagging contraband images. In order to justify the significant authority to coordinate the

[9] *Available at* https://web.archive.org/web/20130921055218/http://www.microsoft.com/global/en-us/news/publishingimages/ImageGallery/Images/Infographics/PhotoDNA/flowchart_photodna_Web.jpg (last accessed Sept. 21, 2013).

scanning and automatic reporting of all images uploaded to Google and other similarly-situated service providers that handle millions of users' private data, the Government should be required to disclose the underlying evidentiary techniques and to show that they are valid and reliable. The Government cannot show that a warrantless search of private files is reasonable where it relies on a private company's matching algorithm to routinely scan millions of files in an attempt to identify contraband images, and it does not provide evidence about the technique or establish the reliability and accuracy of the algorithm.

II. The use of hash algorithms, like other investigative techniques, requires research, testing, and data indicating reliability.

A. On the record before this court, the Government cannot establish with "virtual certainty" that the files it searched were identical to the files that a Google employee previously viewed.

The lower court discussed the concept of a hashing technique, citing a 2005 law review article,[10] but failed to recognize that the file hashing techniques discussed in that article are fundamentally different from image hashing techniques. The court's conclusion turned on the idea that a hash value is equivalent to a "digital fingerprint" and is "uniquely associated with the input data." Mem. Order 2. The Government has not disclosed or even described the Google image matching algorithm, and has not established that it is accurate and

[10] Salgado, *supra*, at 38–39. Mr. Salgado is an attorney and was at that time a senior legal director at Yahoo!. He is now Google's Director of information security and law enforcement matters.

reliable for this function. Without information about the technique, the court had no way to assess its validity or reliability.

The file hashing techniques described in the Salgado law review article are used to uniquely identify or authenticate files and signatures; the image hashing techniques commonly used by Microsoft and others are used to identify similar features in image files even if those files are actually different (e.g. if the color, orientation, or size, has been changed). Microsoft, Digital Crimes Unit, *PhotoDNA* at 4 [hereinafter Microsoft PhotoDNA Slides].[11] While file-hashing algorithms are good at achieving a near-zero percentage of false positive matches, files that have been modified or altered will necessarily produce different hash values. Bruce Schneier, *Applied Cryptography* 30 (1996) ("A single bit change in the pre-image changes, on the average, half of the bits in the hash value."). In contrast, image hashing algorithms provide a way to match images even if they have been altered slightly, but also enable by design the matching of files that do not have the same file-hash values. *See* Microsoft PhotoDNA Slides, *supra*, at 4.

The technique of matching files relies on "one-way hash functions," which are commonly used in cryptographic systems. Schneier, *supra*, at 30. A hash function produces a message digest, which "distill[s] the information contained in

[11] *Available at* https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016802f249e (last accessed Oct. 17, 2018).

a file (small or large) into a single large number, typically between 128 and 256 bits in length." Simson Garfinkel & Gene Spafford, *Web Security & Commerce* 202 (1997). Several message digest algorithms, including MD4 and MD5, were developed by Ronald Rivest,[12] while others (Secure Hash Algorithm, or SHA, and its revised version) were developed by the National Security Agency. *Id.* at 203–204. These functions are "powerful tools for detecting very small changes in very large files." *Id*. at 205. As cryptographer Bruce Schneier explains:

> Think of it as a way of fingerprinting files. If you want to verify that someone has a particular file (that you also have), but you don't want him to send it to you, then ask him for the hash value. If he sends you the correct hash value, then it is almost certain that he has that file.

Schneier, *supra*, at 31.

In contrast, image matching techniques are based on different functions and achieve different results. An image hash is a "distinctive signature which represents the visual content of the image in a compact way (usually just a few bytes)." Sebastiano Battiato, Giovanni Maria Farinella, Enrico Messina, & Giovanni Puglisi, *A Robust Forensic Hash Component for Image Alignment*, 2011 Int'l Conf Image Analysis and Processing 473, 474 (2011). There are many different image hashing techniques because each algorithm is designed to be "robust against allowed operations" while "at the same time" attempting to distinguish different and/or tampered images. *Id*. These "approximate matching"

[12] *See Ron Rivest, The MD5 Message-Digest Algorithm RFC 1321* (Apr. 1992).

techniques are referred to as "perceptual hashing" because they aim to "detect objects that are perceptually similar from the perspective of a human." Petter Christian Bjelland, Katrin Franke, & André Årnes, *Practical Use of Approximate Hash Based Matching in Digital Investigations*, 11 Digital Investigations S18, S20 (2014).

For example, PhotoDNA—a hash function developed by Microsoft and Dartmouth College for use by NCMEC—can match images despite minor changes made to an image that would affect its hash value, such as cropping, resizing, and color adjustments. As Microsoft described at the time that it developed the PhotoDNA technique, "The PhotoDNA 'robust hashing' technique differs from other common hashing technologies because it does not require the image's characteristics to be completely identical to reliably find matches, thereby enabling matches to be identified even when photos are resized or similarly altered." Microsoft, *PhotoDNA: Fact Sheet* (2009).[13] One of the reasons that Microsoft itself cites for the use of its image matching algorithm is that it is capable of matching two images even if the files themselves are different. *See* Microsoft PhotoDNA Slides, *supra*, at 4.

[13] *Available at* https://web.archive.org/web/20140323033617/http://www.microsoft.com/en-us/news/presskits/photodna/docs/photodnafs.doc.

Given the differences in the reliability of file hashing techniques and image hashing techniques, courts must require more information about a hashing method than surface-level assertions and analogies to other forensic techniques. Without this information, it is impossible to determine whether there was a "virtual certainty" that Google staff previously viewed the image files sent to NCMEC.

B. The National Academy of Sciences and other experts have raised significant concerns about the lack of reliable standards for investigative techniques.

EPIC's concerns about the courts' reliance on image matching algorithms in this and other similar cases arise in the context of a growing scientific and legal consensus about the need to assess the reliability and impact of new investigative techniques. Forensic science has been widely criticized because of a lack of clear standards and credible research to support technical conclusions. *See* President's Council of Advisors on Science and Technology, *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods* (2016) [hereinafter PCAST Report]; National Research Council of the National Academies, *Strengthening Forensic Science in the United States: A Path Forward* (2009) [hereinafter National Academy Report]. Even groundbreaking new methods that seem infallible should be subject to scrutiny. Erin E. Murphy, *Inside the Cell: The Dark Side of Forensic DNA*, at x–xi (2015). If the Government believes that Google's algorithm is reliable enough to meet the Fourth Amendment "virtual

certainty" standard, then it should have no problem producing evidence of how the algorithm works and establishing its reliability.

But in many other contexts throughout the criminal justice system, techniques that are presented as infallible or reliable are in fact flawed and imperfect. The 2009 National Academy Report identified several significant problems in forensic science, including "the potential danger of giving undue weight to evidence and testimony derived from imperfect testing and analysis" and the subsequent "admission of erroneous or misleading evidence." National Academy Report at 4. The National Academy Report was commissioned by Congress to "identify the needs of the forensic science community." *See* The Science, State, Justice, Commerce, and Related Agencies Appropriations Act of 2006. P.L. No. 109-108, 119 Stat. 2290 (2005). The expert panel reviewed current forensic methods and made recommendations to help establish guidelines and best practices. National Academy Report at 2. The panel focused on the importance of minimizing the forensic community's "current fragmentation and inconsistent practices," including a lack of "uniformity in certification of forensic practitioners." *Id*. at 6. The Supreme Court has recognized the significance of the National Academy Report in identifying problems with the reliability of forensic methods. *See Melendez-Diaz v. Massachusetts*, 557 U.S. 305, 318 (2009).[14] This

[14] In full, the Court stated:

Court should also look to that report when considering what evidence is necessary to establish reliability of the image matching algorithm at issue in this case, and in other cases involving hash algorithms going forward.

The Supreme Court has recognized that, in the context of the Federal Rules of Evidence, a "trial judge must ensure that any and all scientific testimony or evidence admitted is not only relevant, but reliable." *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579, 589 (1993). The focus of a trial judge should be solely on "principles and methodology . . . ." *Id*. at 595. This presents a problem where "[f]orensic science facilities exhibit wide variability in capacity, oversight, staffing, certification, and accreditation across federal and state jurisdictions." National Academy Report at 14. The National Academy Report made several recommendations for improving the current, fragmented system. Chief among

---

> Nor is it evident that what respondent calls "neutral scientific testing" is as neutral or as reliable as respondent suggests. Forensic evidence is not uniquely immune from the risk of manipulation. According to a recent study conducted under the auspices of the National Academy of Sciences, "[t]he majority of [laboratories producing forensic evidence] are administered by law enforcement agencies, such as police departments, where the laboratory administrator reports to the head of the agency." National Research Council of the National Academies, Strengthening Forensic Science in the United States: A Path Forward 183 (2009) (hereinafter National Academy Report). And "[b]ecause forensic scientists often are driven in their work by a need to answer a particular question related to the issues of a particular case, they sometimes face pressure to sacrifice appropriate methodology for the sake of expediency." *Id.*, at 23–24. A forensic analyst responding to a request from a law enforcement official may feel pressure--or have an incentive--to alter the evidence in a manner favorable to the prosecution.

Melendez-Diaz, 557 U.S. at 318.

them was the establishment and funding of "an independent federal entity, the National Institute of Forensic Sciences ('NIFS')." *Id*. at 19. The Report recommended that NIFS have an advisory board comprised of experts in "forensic science disciplines . . . information technology, measurements and standards, testing and evaluation, law, [and] national security . . . ."*Id*. The NIFS would be responsible for implementing standardized reporting, increasing research, developing best practices, and imposing quality control. *Id*. at 19–33.

A 2016 report from the President's Council of Advisors on Science and Technology ("PCAST"), which sought to clarify the scientific standards underlying the evidentiary rules established in *Daubert* and Rule 702, extended many of the conclusions from the National Academy Report. It noted that "answering the question of scientific validity in the forensic disciplines is important not just for the courts but also because it sets quality standards that ripple out throughout these disciplines—affecting practice and defining necessary research." PCAST Report at 43. The report described the requirement that evidence be based on "reliable principles and methods" to correspond to the scientific standard of "foundational validity." *Id*. Foundational validity requires that, "based on empirical studies," a method be "repeatable, reproducible, and accurate, at levels that have been measured and are appropriate to the intended

application." *Id*. at 47. The report provided the following definitions of repeatable, reproducible, accurate, and reliable:

> By "repeatable," we mean that, with known probability, an examiner obtains the same result, when analyzing samples from the same sources.
> By "reproducible," we mean that, with known probability, different examiners obtain the same result, when analyzing the same samples.
> By "accurate," we mean that, with known probabilities, an examiner obtains correct results both (1) for samples from the same source (true positives) and (2) for samples from different sources (true negatives).
> By "reliability," we mean repeatability, reproducibility, and accuracy.

Id. The report stressed that "[t]he method need not be perfect, but it is clearly essential that its accuracy has been measured based on appropriate empirical testing and is high enough to be appropriate to the application." *Id*. at 48. PCAST made clear that mere assertions of certainty are insufficient: "Statements claiming or implying greater certainty than demonstrated by empirical evidence are scientifically invalid." *Id*. at 54.

A group of law professors, academic researchers, and practicing forensic scientists, led by Dean Jennifer Mnookin, have also sought to develop a common framework for modern forensics. *See* Jennifer L. Mnookin et al., *The Need for a Research Culture in the Forensic Sciences*, 58 UCLA L. Rev. 725 (2011). Dean Mnookin's study argues for an increased focus on empiricism, transparency, and the type of ongoing critical perspective inherent in a "research culture." *Id*. at 740-44. In this case, the Government has the ability to produce evidence describing the

image matching algorithm that it is relying upon and establishing the accuracy and reliability of that technique; but it has failed to do so.

C. The Government's prior use of unreliable techniques to scan and collect private messages underscores the need for proof of reliability.

This is not the first time that the government has purported to develop a technique that perfectly identifies evidence that falls outside the ambit of the Fourth Amendment. In the late 1990s, the FBI developed a software program called "Carnivore" to enable interception of Internet communications pursuant to a court order. *See Internet and Data Interception Capabilities Developed by FBI: Hearing Before the Subcomm. on the Constitution of the H. Comm. on the Judiciary*, 106th Cong. (2000) (statement of Donald M. Kerr, Assistant Director, Laboratory Division, Federal Bureau of Investigation). Carnivore was designed to act like a commercial packet "sniffer" product, which analyzes electronic communications packets as they travel through a network. *See id*. According to the agency, Carnivore could be configured to filter and then store "transmissions which comply with pen register court orders, trap & trace court orders, Title III interception orders, etc." *Id.* The Bureau claimed that, using this technique, only the communications subject to warrant authority would be obtained from the networks of private communications services.

The IIT Research Institute conducted an independent assessment of the FBI's program, and determined that the Carnivore software was capable of collecting "everything that passes by on the Ethernet segment to which it is connected." IIT Research Inst., *Independent Technical Review of the Carnivore System: Final Report* 4-3 (2000) [hereinafter IITRI Final Report]. The Report also found that "Carnivore version 1.3.4 collects more than would be permitted by the strictest possible construction of the pen-trap statute," and the FBI "admitted that a previous version of Carnivore handled pipelined SMTP [packets] incorrectly." *Id*. However, the Report concluded that there were "significant procedural checks to minimize configuration errors." *Id*.

The proper configuration and use of the Carnivore software was thus a critical element of any legal use of the tool. *See Melendez-Diaz,* 557 U.S. at 318. As Professor Orin Kerr also noted, "legitimate concerns exist that the program may malfunction, and as with any tool, human error can cause the program to be configured incorrectly." Orin Kerr, *Internet Surveillance Law After the USA PATRIOT Act: The Big Brother That Isn't*, 97 Nw. U. L. Rev. 607, 654 (2003). In response to this concern, Congress added new reporting requirements under the pen register statute, codified at 18 U.S.C. § 3123(a)(3), that require documentation of:

> (i) any officer or officers who installed the device and any officer or officers who accessed the device to obtain information from the network;
> (ii) the date and time the device was installed, the date and time the device was uninstalled, and the date, time, and duration of each time the device is accessed to obtain information;
> (iii) the configuration of the device at the time of its installation and any subsequent modification thereof; and
> (iv) any information which has been collected by the device

18 U.S.C. § 3123(a)(3).

Without detailed information about the configuration or capabilities of a particular investigative technique, a court cannot determine whether it meets the standard of accuracy and reliability that the Government must establish under the Fourth Amendment.

* * *

Given the high bar established for the private search exception, this Court should require, in addition to information about the operation of Google's hash function, information about the method's accuracy rate, particularly its false positive rate. The consequence of not doing so risks encouraging unreliable law enforcement techniques and weakening constitutional privacy protections, as searches will occur regardless of whether evidence is found. In the absence of such information to establish the validity of the search technique, a search must be deemed unreasonable.

CONCLUSION

This Court should recognize the significant risks posed by the continuous scanning of all images uploaded to Google and other major service providers and the need for transparency and accountability regarding the use of image matching algorithms for criminal investigatory purposes.

October 17, 2018

Respectfully submitted,

/s/ Marc Rotenberg
Marc Rotenberg
Alan Butler
Electronic Privacy Information Center
1718 Connecticut Ave. NW
Suite 200
Washington, DC 20009
(202) 483-1140
Counsel for Amicus

CERTIFICATE OF COMPLIANCE

This brief complies with the type-volume limitation of Fed. R. App. P. 29(d) and Fed. R. App. P. 32(a)(7)(B) because it contains 4,840 words, excluding the parts of the brief exempted by Fed. R. App. P. 32(f) and 6 Cir. R. 32(b)(1). This brief also complies with the typeface requirements of Fed. R. App. P. 32(a)(5) and the type style requirements of Fed. R. App. P. 32(a)(6).

Dated: October 17, 2018

/s/ Alan Butler
Alan Butler

CERTIFICATE OF SERVICE

I hereby certify that on October 17, 2018, I electronically filed the foregoing Brief of *Amicus Curiae* Electronic Privacy Information Center in Support of Appellant with the Clerk of the United States Court of Appeals for the Sixth Circuit using the CM/ECF system. All parties are to this case will be served via the CM/ECF system.

Dated: October 17, 2018

/s/ Alan Butler
Alan Butler