

THE GOOD, THE BAD, AND THE UGLY: BLACK LIVES MATTER PROTESTS, THE JANUARY 6TH INSURRECTION, AND FACIAL RECOGNITION TECHNOLOGY AS ADMISSIBLE EVIDENCE

LAURA C. POWELL*

The debate surrounding law enforcement's use of facial recognition technology (FRT) continues to raise concerns about accuracy, reliability, and equity. Nonetheless, law enforcement agencies continue to purchase, implement, and use FRT as an investigative tool to identify suspects and make arrests. Yet, its ability to enter the courtroom remains untested. Consequently, this Comment seeks to answer the question if FRT was presented as evidence, would it be admissible? To answer that question, this Comment uses the arrests following Black Lives Matter (BLM) protests and the January 6th insurrection. These two events provide contrasting FRT software algorithms, racial compositions, and law enforcement jurisdictions which turn admissibility.

First, the Comment delves into the science of FRT and its development, regulation, and use by law enforcement. Next, the Comment examines the Daubert standard for expert evidence and the Federal Rules of Evidence rules 702 and 403, which a court will use to evaluate FRT. Then with the groundwork laid, the Comment uses the BLM protests and the January 6th insurrection to argue that demographic differences, specific law enforcement policies and training, and the development and testing of the different FRT

* Executive Editor, *American University Law Review*, Volume 72; J.D. Candidate, May 2024, *American University Washington College of Law*, M.B.A., 2019, *University of Delaware*, B.S.M.E., 2008, *University of Delaware*. This would not have been possible without the unwavering support of my parents, John & Robin Elliott and Cleo Powell. My deepest gratitude to James Elliott and Milton Powell Jr. for never letting me forget the importance of my voice. Finally, to Quinn, Celine, and Trace, you all give me perspective and purpose.

software programs will turn admissibility. The Comment concludes that FRT is admissible evidence in the case of January 6th but not BLM because of these differences. This unequal result displays the need for legislative solutions addressing FRT's use and regulation to ensure consistent treatment in the future.

TABLE OF CONTENTS

Table of Contents	278
Introduction.....	279
I. Background	283
A. Facial Recognition Technology	283
1. The science and process of FRT.....	285
2. Law enforcement's use of FRT	286
3. Regulation of FRT.....	288
4. FRT's accuracy: error rates and bias	292
a. System-level error rates & bias.....	293
b. Human interaction & bias.....	297
B. Admissibility Standards for FRT Evidence	300
1. The distinction between Layperson and Expert	300
2. Admissibility standard for expert testimony.	302
a. Testing of the theory or technique.....	304
b. Peer review and publication.....	306
c. Rates of error & controlling standards.....	307
d. General acceptance	309
e. FRE Rule 702: relevancy & fit.....	310
3. Determining admissibility: balancing the Daubert factors	312
II. Analysis	314
A. Testing of the Theory or Technique	317
B. Peer Review and Publication	319
C. Rates of Error & Controlling Standards.....	321
1. Error rates.....	321
2. Controlling standards	324
D. General Acceptance	327
E. Rule 702: Relevancy & Fit	328
F. The Admissibility of FRT for January 6th and BLM.....	330
Conclusion	334

INTRODUCTION

The summer of 2020 was like no other or maybe like those America's collective consciousness hoped were a thing of the past.¹ Restless Americans emerged from COVID lockdowns hoping fresh air would bring freedom. But there was no rest for the weary because a cellphone video on Facebook graphically displaying the murder of George Floyd sparked the restless nation.² That spark landed on the tinder of the many Black bodies already lost at the hands of police. Nationwide, Black Lives Matter ("BLM") protests ensued, denouncing discriminatory policing practices and violence against Black bodies.³ Americans witnessed these protests through cable news networks and cell phone videos uploaded on social media sites.⁴ Yet, law enforcement used the protest video for something else. In fact, the protests illuminated law enforcement's use of facial recognition technology ("FRT") on protestors exercising their First Amendment rights.⁵ In New York City, the New York Police Department (NYPD) descended on a BLM activist's home without a warrant but armed with police dogs, a helicopter, and a facial recognition match.⁶ Similarly, in

1. Gabriel H. Sanchez, *This is What 100 Years of Protests for Racial Justice Looks like in America*, BUZZFEED NEWS (June 5, 2020, 11:34 AM) <https://www.buzzfeednews.com/article/gabrielsanchez/100-years-of-racial-justice-protest-in-america> [https://perma.cc/LFW6-BS35]; Sanya Mansoor, *93% of Black Lives Matter Protests Have Been Peaceful*, TIME (Sept. 5, 2020, 11:47 AM), <https://time.com/5886348/report-peaceful-protests/> [https://perma.cc/5CSG-MEDJ]. For a historical perspective on Black people protesting for equal rights, see *American Experience: Freedom Summer* (PBS June 24, 2014), and Farrell Evans, *The 1967 Riots: When Outrage over Racial Injustice Boiled over*, HISTORY (June 21, 2021), <https://www.history.com/news/1967-summer-riots-detroit-newark-kerner-commission> [https://perma.cc/NTU2-TNC7].

2. See Derrick Bryson Taylor, *George Floyd Protests: A Timeline*, N.Y. TIMES (Nov. 5, 2021), <https://www.nytimes.com/article/george-floyd-protests-timeline.html> [https://perma.cc/KQJ6-DPPL].

3. See *id.*

4. See *id.*

5. Geoffrey A. Fowler, *Protesters Must Keep Big Tech in Check*, WASH. POST, June 14, 2020, at G2; see also Andrew Guthrie Ferguson, *It's Time to Police Surveillance Technology*, WASH. POST, June 28, 2020, at C4 (observing "one of the great ironies of the current protests in the [District of Columbia] . . . is that they are all being monitored by unaccountable federal and local surveillance technologies").

6. Aris Folley, *NYPD Used Facial Recognition Software During Investigation Targeting Black Lives Matter Activist*, HILL (Aug. 19, 2020, 1:15 PM),

Miami, Florida, law enforcement used Clearview AI software to identify and arrest a woman for allegedly throwing a rock at a police officer while participating in a BLM protest.⁷ Similar arrests using different FRT vendors occurred in Washington, D.C.; Columbia, South Carolina; and Philadelphia, Pennsylvania.⁸

Unfortunately, the pain and frustration of 2020 carried into 2021 as America witnessed the January 6th insurrection at the United States Capitol.⁹ This event also aired on television and social media, producing countless photos and videos.¹⁰ In one video, a crowd fought with police over a barricade causing an officer to fall and lose consciousness.¹¹ A man from the group then assaulted two additional officers while gaining access to the restricted area of the Capitol grounds.¹² In this case, the Federal Bureau of Investigation (FBI) used an “open source facial comparison tool, known to provide reliable results in the past” to locate the suspect.¹³ The FBI subsequently arrested and charged him with crimes related to his activities on January 6th.¹⁴ But the FBI was not alone: private citizens also used publicly available FRT systems to identify suspects and provide tips to

<https://thehill.com/policy/technology/512729-nypd-used-facial-recognition-software-during-investigation-targeting-black> [https://perma.cc/2X69-5FYD] (alleging the activist had been matched via facial recognition technology to a video of an assault on an officer with a megaphone, but the felony charges were dropped to a misdemeanor after further investigation).

7. Kate Cox, *Cops in Miami, NYC Arrest Protesters from Facial Recognition Matches*, ARS TECHNICA (Aug. 19, 2020, 4:45 PM), <https://arstechnica.com/tech-policy/2020/08/cops-in-miami-nyc-arrest-protesters-from-facial-recognition-matches/> [https://perma.cc/LLD5-4WQF].

8. *Id.*; Justin Jouvenal & Spencer S. Hsu, *Police Track a Protester's Image*, WASH. POST, Nov. 3, 2020, at B1.

9. Rae Hodge, *Capitol Attack: FBI Mum on Facial Recognition, Clearview AI Searches Spike*, CNET (Jan. 12, 2021, 1:36 PM), <https://www.cnet.com/news/capitol-attack-fbi-mum-on-facial-recognition-clearview-ai-searches-spike/> [https://perma.cc/53AX-4WUR].

10. *Id.*

11. *See* Complaint at 3, *United States v. Randolph*, No. 21-cr-332 (D.D.C. Apr. 16, 2021).

12. *Id.*

13. *Id.* at 7–8 (noting verification occurred using multiple sources, including publicly available YouTube videos and match facial image on a Twitter page).

14. James Vincent, *FBI used Facial Recognition to Identify a Capitol Rioter from his Girlfriend's Instagram Posts*, THE VERGE (Apr. 21, 2021, 6:20 AM), <https://www.theverge.com/2021/4/21/22395323/fbi-facial-recognition-us-capital-riots-tracked-down-suspect> [https://perma.cc/XST2-89RW].

the FBI.¹⁵ Additionally, multiple federal agencies admitted using FRT to the Government Accountability Office (GAO), and Clearview AI touted an increase in searches following January 6th.¹⁶

The attitudes surrounding FRT are as varied as the feelings regarding BLM and January 6th.¹⁷ Supporters defend FRT because it assists law enforcement in bringing unknown assailants to justice for crimes captured on camera.¹⁸ Alternatively, these events intensified concerns regarding system accuracy and the lack of regulation and oversight governing the development and police use of the technology.¹⁹ Detractors decry that inaccurate systems and law enforcement's history of racist, anti-activist surveillance activities raise serious constitutional questions regarding equal justice and other fundamental rights.²⁰

15. See, e.g., Ronan Farrow, *An Air Force Combat Veteran Breached the Senate*, NEW YORKER (Jan. 8, 2021), <https://www.newyorker.com/news/news-desk/an-air-force-combat-veteran-breached-the-senate> [<https://perma.cc/FQB9-ZC9W>] (recounting a private citizen's use of facial recognition and image enhancement to identify and notify the FBI of an insurrectionist on January 6th).

16. U.S. GOV'T ACCOUNTABILITY OFF., GAO-21-518, FACIAL RECOGNITION TECHNOLOGY: FEDERAL LAW ENFORCEMENT AGENCIES SHOULD BETTER ASSESS PRIVACY AND OTHER RISKS 17–18 (2021) (reporting twice as many agencies used FRT on BLM protestors than those investigating January 6th); Vincent, *supra* note 14; Hodge, *supra* note 9.

17. See Transcript of Sentencing at 25–26, *United States v. Mazzocco*, No. 21-cr-00054-TSC (D.D.C. Oct. 4, 2021) (testifying it is a false equivalency to compare BLM protestors to January 6th participants when discussing legal consequences); Gloria Purvis, *The Difference Between Jan. 6 Rioters and Black Lives Matter—and the Catholics Who Refuse to See it*, AM. MAG. (Jan. 6, 2022) <https://www.americamagazine.org/politics-society/2022/01/06/jan-6-capitol-riot-black-lives-matter-242162> [<https://perma.cc/87ZY-HE38>] (discussing the contradiction of those who supported January 6th protestors, but refused to support BLM protests).

18. E.g., James Andrew Lewis, *Regulating Facial Recognition Technology*, CTR. FOR STRATEGIC & INT'L STUD. (June 28, 2021), <https://www.csis.org/analysis/regulating-facial-recognition-technology> [<https://perma.cc/BG39-EJTQ>].

19. See KRISTIN FINKLEA, LAURIE A. HARRIS, ABIGAIL F. KOLKER & JOHN F. SARGENT JR., CONG. RSCH. SERV., R46586, FEDERAL LAW ENFORCEMENT USE OF FACIAL RECOGNITION TECHNOLOGY 8, 15 (2020) [hereinafter R46586].

20. E.g., Shirin Ali, *Communities of Color in NYC Are More Likely to be Surveilled by Facial Recognition Technology*, AMNESTY INTERNATIONAL, HILL (Feb. 15, 2022), <https://thehill.com/changing-america/resilience/smart-cities/594345-communities-of-color-in-nyc-are-more-likely-to-be> [<https://perma.cc/DR5K-NEFP>] (detailing an investigation by Amnesty International on NYPD's disparate use of FRT on communities of color); U.S. GOV'T ACCOUNTABILITY OFF., GAO-21-518, FACIAL

Nevertheless, BLM and January 6th have this in common—despite the lack of transparency, law enforcement used FRT to identify suspects, which led to arrests.²¹ Thus, despite the ongoing dispute over its benefits and the harm it causes, it will become more common as evidence.²² The success of FRT as admissible evidence will depend on the ingenuity of defense attorneys in raising a legal challenge to the reliability and validity of the technology.²³ The use of FRT by law enforcement will inevitably lead to evidence in a criminal trial; still, the question remains—is that acceptable?

RECOGNITION TECHNOLOGY: FEDERAL LAW ENFORCEMENT AGENCIES SHOULD BETTER ASSESS PRIVACY AND OTHER RISKS 17–18 (2021) (reporting twice as many agencies used FRT on BLM protestors than those investigating January 6th); Alex Najibi, *Racial Discrimination in Face Recognition Technology*, HARV. UNIV.: SCI. IN THE NEWS BLOG (Oct. 24, 2020), <https://sitn.hms.harvard.edu/flash/2020/racial-discrimination-in-face-recognition-technology> [<https://perma.cc/3BWY-CCRW>]. Unfortunately, surveillance of Black activists is nothing new but a continuation of historical practices. See Mike German, *The FBI Has a History of Targeting Black Activists. That’s Still True Today*, THE GUARDIAN (June 26, 2020, 5:41 AM) <https://www.theguardian.com/commentisfree/2020/jun/26/fbi-black-activism-protests-history> [<https://perma.cc/DC2W-HAAX>]; Alice Speri, *The FBI Spends a Lot of Time Spying on Black Americans*, THE INTERCEPT (Oct. 29, 2019, 10:07 AM) <https://theintercept.com/2019/10/29/fbi-surveillance-black-activists> [<https://perma.cc/9ZGU-RXWL>].

21. See *supra* notes 13–14, 16; Clare Garvie, Alvaro Bedoya & Jonathan Frankle, *The Perpetual Line-up: Unregulated Police Face Recognition in America*, GEO. L. CTR. PRIV. & TECH. 1, 58–60 (2016), <https://www.perpetuallineup.org> [<https://perma.cc/VZW9-9YDN>]; Mike Holden, *Pittsburgh Police Used Facial Recognition Technology During Black Lives Matter Protests*, WPXI-TV (May 21, 2021, 8:05 AM), <https://www.wpxi.com/news/top-stories/pittsburgh-police-used-facial-recognition-technology-during-black-lives-matter-protests/VT52MGWM3VCDJINJSZPOO5NHKU> [<https://perma.cc/67HS-H2H3>].

22. Cf. Sujit Raman, *Five Principles that Inform the Justice Department’s Use of Facial Recognition Technology*, DOJ (Sept. 15, 2020), <https://www.justice.gov/opa/speech/associate-deputy-attorney-general-sujit-raman-delivers-remarks-community-oriented> [<https://perma.cc/F9VY-7XD2>] (detailing past and continued use of FRT as a tool for effective law enforcement). But see Najibi, *supra* note 20 (positing because of racial discrimination in facial recognition technology it should not be permitted in criminal proceedings).

23. Kaitlin Jackson, *Challenging Facial Recognition Software in Criminal Court*, CHAMPION 14, 14–15 (July 2019). In an interesting turn of events, a public defender used FRT, specifically Clearview AI, to prove a Florida man was innocent. See Kashmir Hill, *Clearview AI, Used by Police to Find Criminals, Is Now in Public Defenders’ Hands*, N.Y. TIMES (Sept. 19, 2022), <https://www.nytimes.com/2022/09/18/technology/facial-recognition-clearview-ai.html> [<https://perma.cc/8C27-BKVW>].

This Comment explores how courts will address the admissibility of facial recognition evidence in criminal trials. Part I provides background on what facial recognition is, how police use it, what regulations and standards exist, and the accuracy and reliability of FRT systems. Then, Part I concludes by detailing the admissibility standards for expert testimony. Next, Part II explores the admissibility of FRT in criminal trials using BLM and January 6th as fact patterns. Finally, the Comment draws together the analysis of FRT in BLM and January 6th, highlighting different factors that will turn a court's analysis and render FRT admissible for January 6th but not for BLM. This Comment concludes courts are equipped to address admissibility in the face of accuracy questions, the lack of regulation and controls, and the dangers of abuse by law enforcement.

I. BACKGROUND

First, Section I.A. will detail the development of FRT and how it works, followed by a discussion on current police use and regulation of the technology. Next, Section I.A. will shift to examine the accuracy of FRT by focusing on how different factors such as age, race, gender, and bias affect error rates. Finally, Section I.B. will discuss the admissibility standards for expert testimony.

A. *Facial Recognition Technology*

FRT resides in the broader field of biometrics, which identifies suspects using anatomical or biological characteristics, often in an automated process.²⁴ The specific physical feature used for identification divides the disciplines within biometrics; fingerprint and iris/retinal scanning are common examples.²⁵ Consequently, identifying people by face falls under facial comparison, which includes both manual and automated methods.²⁶ The brains behind facial comparison are FRT algorithms which automate the analysis and

24. *Overview*, FACIAL IDENTIFICATION SCI. WORKING GRP. (Sept. 19, 2018), https://fiswg.org/FISWG_overview_v2.0_2018_09_19.pdf [https://perma.cc/4T59-FBM2].

25. Susan McCoy, Comment, *O'Big Brother Where Art Thou?: The Constitutional Use of Facial-Recognition Technology*, 20 J. COMPUT. & INFO. L. 471, 473–74 (2002).

26. R46586, *supra* note 19, at 1; *Facial Comparison Overview & Methodology Guidelines*, FACIAL IDENTIFICATION SCI. WORKING GRP. 4 (Oct. 25, 2019), https://fiswg.org/fiswg_facial_comparison_overview_and_methodology_guidelines_V1.0_20191025.pdf [https://perma.cc/L3FT-PKYT].

comparison of faces.²⁷ Most commonly, FRT applications are used for verification and identification.²⁸ Facial verification, also referred to as one-to-one matching, uses a facial image to verify an identity with a database of known individuals.²⁹ Typically, facial verification is used to confirm the identity of a physically present person.³⁰ Common applications are unlocking a phone or accessing customs or immigration checkpoints.³¹

Alternatively, facial identification or one-to-many searching determines the potential identity of an unknown person.³² In order to determine a match, FRT uses an algorithm to analyze a probe image by comparing it to faces in a database. Then, it returns potential match photographs with an associated confidence or similarity score.³³ Referred to as examination, this forensic application requires technical review to verify a match.³⁴ FRT used for identification will be the focus for most future courts because of its use to identify suspects in a forensic application.³⁵

27. U.S. GOV'T ACCOUNTABILITY OFF., GAO-21-526, FACIAL RECOGNITION TECHNOLOGY: CURRENT AND PLANNED USES BY FEDERAL AGENCIES 3–4 (2021); R46586, *supra* note 19, at 2.

28. U.S. GOV'T ACCOUNTABILITY OFF., GAO-21-526, FACIAL RECOGNITION TECHNOLOGY: CURRENT AND PLANNED USES BY FEDERAL AGENCIES 3–4 (2021); Andrew Guthrie Ferguson, *Facial Recognition and the Fourth Amendment*, 105 MINN. L. REV. 1105, 1112–13 (2021) (detailing facial identification, tracking, and verification as applications of facial analysis).

29. Ferguson, *supra* note 28, at 1112. Facial verification can also encompass what is called facial assessment, which is commonly understood as “a quick comparison of image-to-image or image-to-person typically carried out in screening and access control applications.” *Facial Comparison Overview & Methodology Guidelines*, *supra* note 26, at 5.

30. Ferguson, *supra* note 28, at 1113; R46586, *supra* note 19, at 2; PATRICK GROTH, MEI NGAN & KAYEE HANAOKA, NAT'L INST. OF STANDARDS & TECH., NISTIR 8280, FACIAL RECOGNITION VENDOR TEST (FRVT) PART 3: DEMOGRAPHIC EFFECTS I, 4 (2019) [hereinafter NISTIR 8280].

31. *Facial Comparison Overview & Methodology Guidelines*, *supra* note 26, at 4; NISTIR 8280, *supra* note 30, at 1, 4.

32. *Id.* at 3–4; *Overview*, *supra* note 24 (noting facial identification is also referred to as facial review).

33. U.S. GOV'T ACCOUNTABILITY OFF., GAO-21-526, FACIAL RECOGNITION TECHNOLOGY: CURRENT AND PLANNED USES BY FEDERAL AGENCIES 3–4 (2021); *Overview*, *supra* note 24; R46586, *supra* note 19, at 9 & n.46 (noting that similarity scores are often referred to as confidence scores).

34. *Overview*, *supra* note 24.

35. U.S. GOV'T ACCOUNTABILITY OFF., GAO-21-526, FACIAL RECOGNITION TECHNOLOGY: CURRENT AND PLANNED USES BY FEDERAL AGENCIES 6 (2021).

1. *The science and process of FRT*

FRT varies by developer and algorithm, but there is commonality in the underlying science and development process. Generally, to perform facial comparisons the software algorithm measures the size of individual facial features like eyes, nose, and mouth and the distances between them.³⁶ Morphological analysis is the recommended approach for assessing images because it compares two faces by characterizing facial features independently and in relation to the whole face.³⁷ Depending on the developer, algorithms can use other methods for characterizing and matching faces, such as eigenfaces, fisher faces, and principal component analysis.³⁸ Ultimately, the method used by each developer is transparent because the algorithms are considered proprietary and patentable.³⁹

While the calculation methods of algorithms can vary, the steps algorithms execute are generally the same. Most facial recognition programs are machine learning neural networks made of multiple algorithms that detect, characterize, then match facial images.⁴⁰ First, an extraction algorithm detects a facial image in a probe photograph and translates the image into a series of measurements.⁴¹ Then a

36. See Nicholas Bacci, Tobias M. R. Houlton, Nanette Briers & Maryna Steyn, *Validation of Forensic Facial Comparison by Morphological Analysis in Photographic and CCTV Samples*, 135 INT'L J. OF LEGAL MED. 1965, 1967 (2021) (creating a standard guide that provides a feature list to use during morphological analysis, including nose, nostrils, ears, and ear lobes).

37. *Facial Comparison Overview & Methodology Guidelines*, *supra* note 26, at 5–6; see Bacci et al., *supra* note 36, at 1966; Saiba Nazah & Md. Monjurul Islam, *Face Recognition using Morphological Analysis*, 17 GLOB. J. OF COMPUT. SCI. & TECH. 17 (2017) (stating that for facial recognition, linear discriminant analysis and principal component analysis “cannot provide reliable and robust solutions” with complex facial variations).

38. Nazah & Islam, *supra* note 37, at 17; Peter N. Belhumeur, João P. Hespanha & David J. Kriegman, *Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection* 19 IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE 711 (1997); McCoy, *supra* note 25, at 473; see, e.g., *Facial Comparison Overview & Methodology Guidelines*, *supra* note 26, at 5–8 (detailing background on the evolution of FRT methods).

39. See, e.g., NISTIR 8280., *supra* note 30, at 71 (listing the patent number of an algorithm); Facial Recognition, U.S. Patent No. 8,798,336 (filed Sept. 23, 2013); *Clearview AI Principles*, CLEARVIEW AI, <https://www.clearview.ai/principles> [<https://perma.cc/PN65-CLEE>] (“Clearview AI’s proprietary technology processes the image and returns links to publicly available images that contain faces similar to the person pictured in the uploaded image.”).

40. NISTIR 8280, *supra* note 30, at 4.

41. *Id.*

recognition algorithm uses the measures to compare photos and calculate similarity scores.⁴² The algorithm compares the similarity scores against a threshold value and provides it to the user if it is within the acceptable range (i.e., passes the threshold set); if not, it is rejected.⁴³ Finally, a human reviewer confirms a mate or imposter based on the system's recommendations.⁴⁴ Images of the same person are considered a positive match or "mate," but different people result in a "nonmate" or imposter.⁴⁵ Errors occur when an imposter score is above, or a mate score is below the given threshold; this results in false positives and negatives, respectively.⁴⁶ When the algorithm fails to match two facial images of the same person accurately, it is a false negative.⁴⁷ Alternatively, it is a false positive when the algorithm incorrectly matches two facial images of different people.⁴⁸

2. *Law enforcement's use of FRT*

Law enforcement uses FRT to investigate and identify suspects, locate missing persons, expose identity theft, and generally solve crimes.⁴⁹ Typically, law enforcement obtains a probe photo from a surveillance video and uses FRT to obtain an identification.⁵⁰ When the FRT program returns potential matches, law enforcement compares

42. *Id.*

43. R46586, *supra* note 19, at 9.

44. U.S. GOV'T ACCOUNTABILITY OFF., GAO-21-526, FACIAL RECOGNITION TECHNOLOGY: CURRENT AND PLANNED USES BY FEDERAL AGENCIES 13 (2021); PATRICK GROTH, MEI NGAN & KAYEE HANAOKA, NAT'L INST. OF STANDARDS & TECH., NISTIR 8238, ONGOING FACE RECOGNITION VENDOR TEST (FRVT) PART 2: IDENTIFICATION 12 (2018) [hereinafter NISTIR 8238].

45. NISTIR 8280, *supra* note 30, at 4.

46. R46586, *supra* note 19, at 9–10.

47. *Id.* at 9.

48. *Id.*

49. See Associate Deputy Attorney General Sujit Raman Delivers Remarks at the Community Oriented Policing Services (COPS)/Police Executive Research Forum (PERF) Facial Recognition Technology Forum, DOJ (Sept. 15, 2020), <https://www.justice.gov/opa/speech/associate-deputy-attorney-general-sujit-raman-delivers-remarks-community-oriented> [<https://perma.cc/Z27L-4SWF>] (detailing past and continued use of FRT as a tool for effective law enforcement); McCoy, *supra* note 25, at 473 (stating how this technology is used for security purposes).

50. *Facial Recognition Technology (Part II): Ensuring Transparency in Government Use: Hearing Before the H. Comm. on Oversight & Reform*, 116th Cong. 41 (2019) [hereinafter *Hearings*] (statement of Kimberly J. Del Greco, Deputy Assistant Director, FBI); see *New York v. Reyes*, 69 Misc. 963, 964 (N.Y. Sup. Ct. 2020) (using surveillance footage and FRT software, law enforcement identified a burglary suspect).

them with the source material to confirm a positive match.⁵¹ In some cases, the matches are part of a lineup presented to an eyewitness for identification.⁵²

The FBI has used FRT since 2011⁵³ and operates two systems to perform searches.⁵⁴ First, the FBI maintains and operates the Next Generation Identification-Interstate Photo System (NGI-IPS), which supports state and local law enforcement by allowing officials to submit photos for matching.⁵⁵ The NGI-IPS database contains over 93 million photos built from criminal mugshots, photos submitted for criminal background checks, and photos of unknown suspects believed to have committed a felony.⁵⁶

Similarly, the Facial Analysis, Comparison, and Evaluation (FACE) Services Unit supports matching solely for the FBI.⁵⁷ The FACE Services Unit searches probe photos against other state and federal systems, such as driver's license, visa, and passport databases, in addition to the NGI-IPS database.⁵⁸ The results are provided to an FBI image specialist for review and confirmation, who then returns no more than two photos to the requesting agent.⁵⁹ However, the FBI does not verify the accuracy of external FRT systems because the state or other federal agencies own and operate the systems.⁶⁰

The FBI asserts that FRT is a tool used in conjunction with a valid criminal investigation to produce potential leads and expressly prohibits FRT results for positive identification without

51. *Hearings, supra* note 50, at 4 (statement of Kimberly J. Del Greco, Deputy Assistant Director, FBI); *see Reyes*, 69 Misc. at 964 (utilizing the FRT results, the law enforcement office examined the video again to confirm the match).

52. *See Reyes*, 69 Misc. at 965.

53. U.S. GOV'T ACCOUNTABILITY OFF., GAO-19-579T, FACIAL RECOGNITION TECHNOLOGY: DOJ AND FBI HAVE TAKEN SOME ACTIONS IN RESPONSE TO GAO RECOMMENDATIONS TO ENSURE PRIVACY AND ACCURACY, BUT ADDITIONAL WORK REMAINS 2 (2019) [hereinafter GAO-19-579T] (specifying the FBI began piloting NGI-IPS in 2011 and the system became fully operational in 2015).

54. *Hearings, supra* note 50, at 4 (statement of Kimberly J. Del Greco, Deputy Assistant Director, FBI).

55. R46586, *supra* note 19, at 5; *Hearings, supra* note 50, at 10 (statement of Kimberly J. Del Greco, Deputy Assistant Director, FBI).

56. R46586, *supra* note 19, at 6.

57. *Id.* at 5–6.

58. GAO-19-579T, *supra* note 53, at 3, 5–6 (containing over 641 million facial images, with many collected independent of consent or criminal history).

59. *Id.* at 6 (highlighting no matches are returned in some cases).

60. *Id.* at 17.

corroboration.⁶¹ Additionally, the FBI contends other law enforcement agencies that use their system are held to the same standard.⁶² Even so, the GAO recommended regular operational reviews, but the FBI responded it was unnecessary because it solicited feedback, received no complaints, and found no cases of unauthorized use.⁶³ Unfortunately, these assertions are absent independent evaluations or verification.⁶⁴ Similarly, state and local police departments also claim they only use FRT for investigative purposes.⁶⁵ Nonetheless, there are examples of the NYPD and Detroit Police Department conducting seemingly uncorroborated arrests using FRT.⁶⁶

3. Regulation of FRT

There is no direct federal statutory or regulatory scheme for FRT design, use, or performance; thus, it has developed unregulated.⁶⁷ However, there are two tangentially related statutes that apply to federal agencies' collection and use of personal data; the Privacy Act of 1974⁶⁸ and the E-Government Act of 2002.⁶⁹ These statutes require the FBI and Department of Justice (DOJ) to produce and update publicly available privacy impact assessments and system of records

61. See *Hearings*, *supra* note 50, at 4 (statement of Kimberly J. Del Greco, Deputy Assistant Director, FBI) (explaining that the photos returned are not considered possible matches themselves).

62. R46586, *supra* note 19, at 5.

63. GAO-19-579T, *supra* note 53, at 17.

64. R46586, *supra* note 19, at 15; Garvie et al., *supra* note 21.

65. NYPD PATROL GUIDE: FACIAL RECOGNITION TECHNOLOGY (Mar. 12, 2020); DETROIT POLICE DEP'T, MANUAL DIRECTIVE 307.5 FACIAL RECOGNITION (Sept. 12, 2019).

66. Folley, *supra* note 6; Bobby Allyn, 'The Computer Got It Wrong': How Facial Recognition Led to False Arrest of Black Man, NPR (June 24, 2020, 8:00 AM), <https://www.npr.org/2020/06/24/882683463/the-computer-got-it-wrong-how-facial-recognition-led-to-a-false-arrest-in-michig> [<https://perma.cc/Z2AL-F2WN>].

67. R46586, *supra* note 19, at 15; see also Brad Smith, *Facial Recognition: It's Time for Action*, MICROSOFT: ON THE ISSUES (Dec. 6, 2018), <https://blogs.microsoft.com/on-the-issues/2018/12/06/facial-recognition-its-time-for-action/> [<https://perma.cc/7VML-9QDT>] ("We believe it's important for governments in 2019 to start adopting laws to regulate this technology. The facial recognition genie, so to speak, is just emerging from the bottle."). For example, private companies are free to advertise high accuracy rates, yet specifically disclaim them in contracts with police departments. Garvie et al., *supra* note 21 (detailing contract with San Diego Association of Governments in which FaceFirst purports to be 95% accurate but then disclaims liability for failing to meet that in contracts).

68. 5 U.S.C. § 552a.

69. 44 U.S.C. §§ 3601-06; GAO-19-579T, *supra* note 53, at 6.

notices.⁷⁰ Even so, the FBI and DOJ only recently complied with these laws. A 2016 GAO audit revealed privacy impact statements did not occur until three years after the FRT systems were piloted.⁷¹ Further, there were no systems of records notices at the time of the audit.⁷² The subsequent GAO report in 2018 showed improvement but not full compliance.⁷³

The fact is, there are millions of law-abiding individuals currently in FRT databases.⁷⁴ The FBI and at least twenty-six states have agreements leveraging existing laws, often absent the knowledge or consent of legislators.⁷⁵ Many of the laws used to facilitate these agreements predate FRT and do not require informed consent to include individuals in the databases.⁷⁶ Congress has brought up this deficiency as a potential violation of civil rights.⁷⁷ In response, the FBI touts results of internal audits showing full compliance with policy and procedures, asserting from 2017 to 2019, “there have been no findings of civil

70. GAO-19-579T, *supra* note 53, at 6–7.

71. U.S. GOV'T ACCOUNTABILITY OFF., GAO-16-267, FACIAL RECOGNITION TECHNOLOGY: FBI SHOULD BETTER ENSURE PRIVACY AND ACCURACY 18–19 (2016).

72. *See id.* at 8–9, 12 (highlighting that privacy impact statements were provided three years after pilots and systems of records notices did not occur until after the GAO's audit five years after initiation of the database).

73. *Id.* at 12–13.

74. *Hearings, supra* note 50, at 1–2 (statement of Rep. Elijah Cummings, Chairman, H. Comm. On Oversight & Reform); *see also* GAO-19-579T, *supra* note 53, at 5–6 (reporting the FACE Services Unit has access to over 641 million facial images); KRISTIN FINKLEA & KELSEY SANTAMARIA, CONG. RSCH. SERV., IN11614, U.S. CAPITOL ATTACK AND LAW ENFORCEMENT USE OF FACIAL RECOGNITION TECHNOLOGY 2 (2021) (citing Clearview AI amassing over three billion images to create its search database).

75. *Hearings, supra* note 50, at 2–3 (statement of Rep. Jim Jordan, Ranking Member, H. Comm. On Oversight & Reform); Garvie et al., *supra* note 21 (finding that at least twenty-six states have agreements and one in two Americans are in a searchable database).

76. Garvie et al., *supra* note 21; *see also Hearings, supra* note 50, at 2–3 (statement of Rep. Jim Jordan, Ranking Member, H. Comm. On Oversight & Reform) (noting that over twenty State Bureaus of Motor Vehicles have given individuals' driver license data to the FBI without seeking their consent).

77. *Cf. Hearings, supra* note 50, at 2–3 (statement of Rep. Jim Jordan, Ranking Member, H. Comm. On Oversight & Reform) (explaining the lack of informed consent and a standard of suspicion for use implicates due process concerns); Garvie et al., *supra* note 21 (highlighting how only a few law enforcement agencies have specific policies regarding FRT use during protests).

liberties violations or evidence of system misuse.”⁷⁸ However, a GAO audit showed the results were incomplete because the FBI’s report excluded state and local users of FBI systems as well as the FBI’s use of external systems.⁷⁹ Additionally, there is no mechanism to regulate or enforce misuse in real time.⁸⁰ Similarly, many state and local law enforcement agencies do not audit misuse or have publicly available use policies.⁸¹

While federal regulation is lacking, some states have enacted statutes to protect biometric information and directly regulate private companies.⁸² One of the first was the Illinois Biometric Information Privacy Act⁸³ (“BIPA”). The BIPA allows individuals to control their personal biometric data by prohibiting private companies from collecting it without informed consent.⁸⁴ Since Illinois passed the BIPA in 2008, other states have followed suit.⁸⁵ However, the BIPA and similar laws do not address state or police uses of biometric data; instead, they prohibit private companies and individuals from using such data in certain circumstances.⁸⁶ Similarly, some statutes regulate

78. DEP’T OF JUST., STATEMENT OF KIMBERLY J. DEL GRECO BEFORE THE COMMITTEE ON OVERSIGHT AND REFORM AT A HEARING CONCERNING “THE USE OF FACIAL RECOGNITION TECHNOLOGY BY GOVERNMENT ENTITIES AND THE NEED FOR OVERSIGHT OF GOVERNMENT USE OF THIS TECHNOLOGY UPON CIVILIANS” 3 (2019).

79. GAO-19-579T, *supra* note 53, at 12.

80. *See Hearings, supra* note 50, at 21–22 (statement of Kimberly J. Del Greco, Deputy Assistant Director, FBI) (arguing over the lack of accountability in the FBI system).

81. GAO-19-579T, *supra* note 53, at 12 (concluding of fifty-two agencies surveyed, only four have publicly available policies and nine audit misuse).

82. Garvie et al., *supra* note 21; *see, e.g.*, Biometric Information Privacy Act, 740 ILL. COMP. STAT. 14/15 (2022).

83. 740 ILL. COMP. STAT. 14/15 (West 2008).

84. *Id.*; *see also Biometric Information Privacy Act (BIPA)*, ACLU, <https://www.aclu-il.org/en/campaigns/biometric-information-privacy-act-bipa> [<https://perma.cc/LU9D-FQY5>].

85. TEX. BUS. & COM. § 503.001 (West 2017); WASH. REV. CODE ANN. § 19.375.020 (West 2017); California Consumer Privacy Act of 2018, CAL. CIVIL CODE § 1798.100-199 (West 2018); Stop Hacks and Improve Electronic Data Security (SHIELD) Act, N.Y. GEN. BUS. LAW §§ 899-aa, 899-bb (McKinney 2020).

86. *See, e.g.*, TEX. BUS. & COM. § 503.001 (providing that a “person may not capture a biometric identifier” without prior consent, may not sell biometric data without consent or unless allowed by law); WASH. REV. CODE ANN. § 19.375.020 (prohibiting any company or individual from entering biometric data “in a database for a commercial purpose, without first providing notice, obtaining consent, or providing a

disclosure and implement oversight, not the technology itself or the use of it.⁸⁷

Some states and local municipalities have statutes restricting or banning FRT; however, the bans and regulations have been rolled back as crime rates creep up.⁸⁸ Alabama, Maine, Massachusetts, and Baltimore, Maryland, remain examples of places where police are completely banned from using FRT or require a warrant.⁸⁹ But, Virginia eliminated its FRT prohibition after only a year, and New Orleans removed its ordinance after less than two years, due partly to rising crime rates.⁹⁰ Finally, in some states, lack of regulation remains a concern. For example, in Florida, up to 8,000 searches occur monthly, yet nothing has been disclosed to the public defenders' office, which could serve as a check on the use of this technology for criminal justice purposes.⁹¹

Despite the efforts to standardize FRT and professionalize the workforce using it, the absence of legislation means adherence to standards and best practices is voluntary. The National Institute of Standards and Technology (NIST)⁹² has established a Facial

mechanism to prevent the subsequent use of a biometric identifier for a commercial purpose”).

87. N.Y.C., N.Y., ADMIN. CODE § 14-188 (2020); *see also* Lauren Feiner, *NYC Lawmakers Pass Bill Requiring Police to Disclose Surveillance Technology*, CNBC (June 18, 2020, 6:18 PM), <https://www.cnbc.com/2020/06/18/nyc-passes-bill-requiring-police-to-disclose-surveillance-technology.html> [<https://perma.cc/XWC3-WHAZ>] (detailing the ACLU initiative to get cities to adopt laws increasing transparency about surveillance, like San Francisco, Palo Alto and Oakland, CA as well as Yellow Springs, Ohio, Nashville, Tennessee, and three cities in Massachusetts).

88. Paresh Dave, *U.S. Cities are Backing Off Banning Facial Recognition as Crime Rises*, REUTERS (May 12, 2022, 1:35 PM), <https://www.reuters.com/world/us/us-cities-are-backing-off-banning-facial-recognition-crime-rises-2022-05-12> [<https://perma.cc/C6GY-7WCA>].

89. ALA. CODE § 15-10-110-111 (2022); Baltimore, Md., Ordinance 21-0001 (Aug. 9, 2021); ME. REV. STAT. ANN. tit. 25, § 6001 (West 2021); MASS. GEN. LAWS ANN. ch. 6, § 220 (West 2021).

90. Dave, *supra* note 88; Kate Kaye, *Police Can Use Facial Recognition Again After Ban in New Orleans, Home to Sprawling Surveillance*, PROTOCOL (July 26, 2022), <https://www.protocol.com/enterprise/new-orleans-surveillance-facial-recognition> [<https://perma.cc/2AAZ-KLPP>].

91. Garvie et al., *supra* note 21.

92. NIST is a non-regulatory federal agency within the Department of Commerce that collaborates with federal agencies, law enforcement, industry, and academic partners to conduct research, testing and evaluation, technical guidance and support,

Identification Scientific Working Group (FISWG) to develop FRT standards, guidelines, and best practices.⁹³ FISWG recommends research and development areas to further the face analysis discipline.⁹⁴ Standardization is crucial because untrained individuals can misidentify suspects up to 50% of the time.⁹⁵ Consequently, the FBI requires all users to complete training that meets the FISWG standards and recommends other law enforcement agencies require similar training before reviewing matches.⁹⁶ The NYPD and the Detroit Police Department follow similar rules restricting use for positive identification but do not refer to specialized training in their controlling documents.⁹⁷ Compliance is not a foregone conclusion because Detroit only recently modified its internal procedures after a highly publicized false arrest.⁹⁸ The NYPD has similar instances where arrests appeared to be based on FRT alone in direct contradiction with internal policies.⁹⁹ Ultimately, the lack of a clear policy or regulation makes the use of FRT inconsistent and disclosure voluntary in most locations.

4. *FRT's accuracy: error rates and bias*

The NIST has tested FRT for over ten years, systematically broadening the scope from basic to more specific evaluations, like demographics. The NIST has reported significant gains in accuracy in

and standards related to FRT. *About NIST*, NAT'L INST. OF STANDARDS & TECH., <https://www.nist.gov/about-nist> [<https://perma.cc/4S29-S2KX>].

93. *About Scientific Working Groups*, FACIAL IDENTIFICATION SCI. WORKING GRP., <https://fiswg.org/about.html> [<https://perma.cc/7KCZ-DF9A>].

94. *Id.*; R46586, *supra* note 19, at 3.

95. See Garvie et al., *supra* note 21; see also P. Jonathon Phillips, Amy N. Yates, Ying Hu, Carina A. Hahn, Eilidh Noyes, Kelsey Jackson et al., *Face Recognition Accuracy of Forensic Examiners, Superrecognizers, and Face Recognition Algorithms*, 115 PROC. NAT'L ACAD. SCI. 24, 76 (2018) (concluding the best performing professional face examiners were equal to the best machines but the most accurate results occur through collaboration between trained examiners and precise systems).

96. *Hearings*, *supra* note 50, at 4–5 (statement of Kimberly J. Del Greco, Deputy Assistant Director, FBI); see also *Guide for Role-Based Training in Facial Comparison*, FACIAL IDENTIFICATION SCI. WORKING GRP. (July 17, 2020), <https://fiswg.org/documents.html> [<https://perma.cc/5FTV-P8TE>] (reporting the training level depends on the purpose and the quality, quantity, and complexity of the images to be analyzed).

97. NYPD PATROL GUIDE, *supra* note 65; DETROIT POLICE DEP'T, *supra* note 65.

98. See Allyn, *supra* note 66.

99. *E.g.*, Folley, *supra* note 6.

the last five years, but there are several caveats to consider.¹⁰⁰ First, many algorithms do not achieve sub-1% error rates, and those that do achieve this rate for mugshot images searched in mugshot databases.¹⁰¹ The NIST studies are clear—low image quality decreases accuracy; thus, the study cautions that the low error rates correspond to “mostly excellent cooperative live-capture mugshot images.”¹⁰² Second, the error rate variation across developers is from 0.1% to over 50% for false negatives, which displays a lack of consensus and standardization.¹⁰³ Because the test is performed in a “black box” without access to the source code, there is no way to verify which algorithms use the more accurate “(deep) convolutional neural networks.”¹⁰⁴ Third, the error rate will increase as the database size increases, which is important because the study population size is much smaller than most operational databases.¹⁰⁵ Finally, subject demographics such as race and age can cause error rates to increase.¹⁰⁶ Nonetheless, the reports are some of the only comprehensive evaluations of numerous algorithms by an independent entity.¹⁰⁷

a. System-level error rates & bias

Across all algorithms, accuracy diminishes for low-quality images, those with significant aging, women, and Black and Brown individuals.¹⁰⁸ Taking each in turn—first, low-quality images are affected by resolution, lighting, pose, angle, facial orientation, and

100. NISTIR 8238, *supra* note 44, at 2 (evaluating 127 algorithms and a data set of 32.3 million photographs, both high and low quality and indicating the gain exceeds that of the previous testing period from 2010–2013).

101. *Id.* at 6.

102. PATRICK GROTH, MEI NGAN & KAYEE HANAOKA, NAT’L INST. OF STANDARDS & TECH., NISTIR 8271 (DRAFT SUPPLEMENT), FACE RECOGNITION VENDOR TEST (FRVT) PART 2: IDENTIFICATION 7 (2022) [hereinafter NISTIR 8271].

103. NISTIR 8238, *supra* note 44, at 4–6.

104. *Id.* at 4–5 (detailing the black box test means “the templates that hold features extracted from face images are entirely proprietary opaque binary data that embed considerable intellectual property of the developer”).

105. *Id.* at 7.

106. NISTIR 8280, *supra* note 30, at 2–3.

107. NISTIR 8238, *supra* note 44, at 2 (evaluating 127 algorithms and a dataset of 32.3 million photographs, both high and low quality).

108. *Id.* at 6–7 (concluding error rates can vary from “a few tenths of one percent up to beyond fifty percent”); *see also id.* at 7 (explaining both image quality and aging faces can hurt the accuracy of an algorithm).

obstructions such as hats, masks, and hair.¹⁰⁹ Second, aging or facial image changes over time cause similarity scores to decline, progressively decreasing accuracy.¹¹⁰ Consequently, both image quality and aging result in increased false negatives.¹¹¹ Third, false positives are higher in women, children, and the elderly but lower in middle-aged adults.¹¹² Finally, false-negative rates are higher in Black, Asian, and American Indian people and lowest in Eastern European populations.¹¹³ Race can vary false positives from a factor of 10 to 100 and false negatives by factors below three.¹¹⁴ Ostensibly, this means FRT sees Black and Brown faces as similar when they are, in fact, not.

The reported error rate is only part of the story because most reports and studies do not directly report false negative and positive rates at threshold values.¹¹⁵ Generally, a higher threshold returns fewer potential matches than a lower threshold.¹¹⁶ Specifically, a high threshold increases the number of false negatives, while a lower threshold increases the number of false positives.¹¹⁷ The threshold for most systems is set internally and is not adjusted for different cameras, environmental conditions, or demographics, making it difficult for the user to know if the reported error rate applies to their use and application.¹¹⁸ For example, while the FBI determined its detection rate to be 86%, the GAO concluded this was only valid when a list of fifty candidates was returned and not when fewer matches were requested.¹¹⁹ According to the FBI, smaller lists would lower accuracy (i.e., a higher threshold would lower accuracy).¹²⁰ The GAO recommended further testing to characterize false positives and negatives errors at varying thresholds.¹²¹ Disturbingly, the DOJ

109. *Facial Comparison Overview & Methodology Guidelines*, *supra* note 26, at 6 (stating obstruction of facial features hurts the accuracy of an algorithm).

110. NISTIR 8238, *supra* note 44, at 7 (noting the impact of drug use and face-altering surgery is currently unknown).

111. NISTIR 8280, *supra* note 30, at 7.

112. *E.g.*, *id.* at 8.

113. *See id.* at 1–2.

114. *Id.*

115. *Id.* at 3; *see* R46586, *supra* note 19, at 9.

116. NISTIR 8280, *supra* note 30, at 6.

117. *Id.*

118. *Id.*

119. *See* GAO-19-579T, *supra* note 53, at 14.

120. *Id.*

121. *Id.*

determined NGI-IPS cannot produce false positives because it is not used for positive identification; the GAO flatly disagreed.¹²² Specifically, to provide a complete picture, systems must report false negative and positive rates at the threshold across demographic groups.¹²³ But, most reports and studies do not report errors in this manner; thus, it is an area that requires further evaluation to determine if current technology can effectively mitigate the errors.¹²⁴

Importantly, these factors can combine to compound inaccuracies, making error rates algorithm and fact-dependent. For example, Massachusetts Institute of Technology's Gender Shades study determined the poorest accuracy existed for Black females between 18-30 years old, with error rates up to 34% higher than for lighter-skinned males.¹²⁵ Similarly, the false-negative rate for Black and Caribbean individuals is higher when combined with low-quality images.¹²⁶ Even photographs considered high-quality in some applications can produce lower accuracy when combined with other factors. For example, pictures taken in standardized environments (like mugshot photos) where default camera settings are not optimized for varying skin tones—result in lower accuracy for darker skin.¹²⁷

Similarly, population growth from testing and training databases to real-world applications can increase errors related to aging and race.¹²⁸ The NIST vendor test for identification used a population size of 640,000 mugshot photographs to determine the most accurate

122. *Id.* at 15 (determining the 86% accuracy rate absent a false positive rate is incomplete).

123. NISTIR 8280, *supra* note 30, at 3.

124. *See id.*

125. *Gender Shades*, MIT MEDIA LAB, <https://www.media.mit.edu/projects/gender-shades/overview> [<https://perma.cc/386B-VRBV>]; Najibi, *supra* note 20.

126. NISTIR 8280, *supra* note 30, at 2 (using border crossing photos as an example).

127. *See* Najibi, *supra* note 20. Interestingly, Google also recognized default cameras do not do justice to Black skin and are updating their phone cameras accordingly. *See* Allison Johnson, *Google is Trying to Make Its Image More Inclusive*, THE VERGE (May 18, 2021, 3:34 PM), <https://www.theverge.com/2021/5/18/22442515/google-camera-app-inclusive-image-equity-skintones> [<https://perma.cc/QEE7-4WH3>] (reporting on Google's announcement that it will be making photos of people of color more accurate).

128. NISTIR 8280, *supra* note 30, at 1, 10 (analyzing the accuracy of face recognition vendor tests across demographic groups based on four data sets collected from U.S. government applications and recommending more diverse training data to mitigate false positives).

algorithm error rate was 0.27%.¹²⁹ When the database size increased to 12 million, the error rate grew to 0.45%.¹³⁰ When aging was factored on top, the error rate increased by 1% to more than 10%, depending on the algorithm.¹³¹ Similarly, when the NIST performed the demographic vendor test with a dataset of approximately 18 million images focusing on demographic differences, the error rates were between 0.5% to above 10%.¹³² These error rates are not using images from video surveillance.¹³³ It is unknown what the magnitude of increase is for the FBI database containing upwards of 93 million photos.¹³⁴

Additionally, accuracy declines if developers do not use diverse data sets when training algorithms.¹³⁵ The training data set should have numerous known unique identities with multiple images, and a wide range of poses, lighting, and demographic differences.¹³⁶ Unfortunately, standard training databases have been predominately white and male, contributing to the disparate accuracy rate across demographics.¹³⁷ Developers like IBM and Microsoft acknowledged this bias and reduced inaccuracies by modifying testing, enhancing

129. NISTIR 8238, *supra* note 44, at 41.

130. *Id.* at 6.

131. *Id.*

132. *Id.* at 7.

133. *Id.*; R46586, *supra* note 19, at 6.

134. NISTIR 8238, *supra* note 44, at 7 (discussing how accuracy decreases as population size increases).

135. FACIAL IDENTIFICATION SCI. WORKING GRP., FACE RECOGNITION SYSTEMS OPERATION ASSURANCE: IDENTITY GROUND TRUTH 4 (2021) [hereinafter FISWG IDENTITY GROUND TRUTH]; see also NISTIR 8280, *supra* note 30, at 7, 9–10 (noting lower rates of false positives where training data demographics match operational data demographics).

136. FISWG IDENTITY GROUND TRUTH, *supra* note 135, at 4.

137. Najibi, *supra* note 20; Steve Lohr, *Facial Recognition is Accurate, If You're a White Guy*, N.Y. TIMES (Feb. 9, 2018), <https://www.nytimes.com/2018/02/09/technology/facial-recognition-race-artificial-intelligence.html> [<https://perma.cc/CMS9-SP98>]; William Crumpler, *The Problem of Bias in Facial Recognition*, CTR. FOR STRATEGIC & INT'L STUDS: STRATEGIC TECHS. BLOG (May 1, 2020), <https://www.csis.org/blogs/technology-policy-blog/problem-bias-facial-recognition> [<https://perma.cc/PY6T-NQL7>]; Ali Breland, *How White Engineers Built Racist Code – and Why it's Dangerous for Black People*, THE GUARDIAN (Dec. 4, 2017, 4:00 AM), <https://www.theguardian.com/technology/2017/dec/04/racist-facial-recognition-white-coders-black-people-police> [<https://perma.cc/E8ZD-G6LW>].

data collection diversity, and sometimes halting development.¹³⁸ When the American Civil Liberties Union assessed Amazon’s software finding similar inaccuracies and bias, Amazon’s response did not address racial bias but attacked the evaluation.¹³⁹

b. Human interaction & bias

Scientific evidence, like FRT, involves individual decision-making; thus, it is not immune from cognitive bias, which can increase human errors and contribute to wrongful convictions.¹⁴⁰ Cognitive bias occurs because the brain uses established processes to accomplish efficient decision-making through a “top-down” approach utilizing knowledge, experience, motivations, and emotional states to comprehend data.¹⁴¹ A common form of cognitive bias is confirmation bias, which occurs

138. *IBM Abandons ‘Biased’ Facial Recognition Technology*, BBC NEWS (June 9, 2020), <https://www.bbc.com/news/technology-52978191> [<https://perma.cc/2SQF-HYT4>]; Shaun Sutner, *Microsoft Reins in AI Facial and Voice Recognition Tech*, TECHTARGET (June 23, 2022) <https://www.techtargget.com/searchenterpriseai/news/252521952/Microsoft-reins-in-AI-facial-and-voice-recognition-tech> [<https://perma.cc/2LQY-XJ4G>]. *But see* Kyle Wiggers, *Bias Persists in Face Detection Systems from Amazon, Microsoft, and Google*, VENTURE BEAT (Sept. 3, 2021, 10:40 AM), <https://venturebeat.com/ai/bias-persists-in-face-detection-systems-from-amazon-microsoft-and-google> [<https://perma.cc/SL6P-YSNL>] (highlighting a study which found despite corporate acknowledgment of discrimination issues within their software, facial and voice recognition technologies by the companies continue to present obvious biases against certain races, genders, and ethnicities).

139. Jacob Snow, *Amazon’s Face Recognition Falsely Matched 28 Members of Congress with Mugshots*, ACLU (July 26, 2018), <https://www.aclu.org/news/privacy-technology/amazons-face-recognition-falsely-matched-28> [<https://perma.cc/T3Z5-YBC8>]; Matt Wood, *Thoughts on Machine Learning Accuracy*, AMAZON WEB SERVS.: AWS NEWS BLOG (July 27, 2018), <https://aws.amazon.com/blogs/aws/thoughts-on-machine-learning-accuracy/> [<https://perma.cc/A876-FKKN>].

140. Sherry Nakhaeizadeh, Itiel E. Dror & Ruth M. Morgan, *The Emergence of Cognitive Bias in Forensic Science and Criminal Investigations*, 4 BRIT. J. AM. LEGAL STUD. 527, 528–29 (2015); D. Kim Rossmo & Joycelyn M. Pollock, *Confirmation Bias and Other Systemic Causes of Wrongful Convictions: A Sentinel Events Perspective*, 11 NE. U. L. REV. 790, 802 (2019) (discussing a study on how wrongful convictions occur). *See generally* NAT’L. RSCH. COUNCIL, STRENGTHENING FORENSIC SCIENCE IN THE UNITED STATES: A PATH FORWARD 1, 8 (2009) (noting research into minimizing selective interpretation and cognitive bias in forensic science is still largely needed).

141. Nakhaeizadeh et al., *supra* note 140, at 535–36 (noting this process is similar to heuristics, “mental shortcuts in decision making” that sacrifice accuracy for speed and less effort); *see also* Rossmo & Pollock, *supra* note 140, at 811 (explaining confirmation bias encompasses both “cognitive processes (limited ability to handle complex tasks) and motivated processes (influence of desire on belief, consistency needs)”).

when the brain looks through memories and information to confirm pre-existing beliefs to be more efficient.¹⁴² A recent study evaluated fifty wrongful convictions and determined that confirmation bias was present in 74% of all cases and was the leading causal factor.¹⁴³ Similarly, tunnel vision, which obstructs the accuracy of perception and interpretation (a result of cognitive bias), was the second-highest causal factor.¹⁴⁴ This is important because criminal investigations operate off a theory; thus, they are particularly susceptible to forms of cognitive bias, such as tunnel vision.¹⁴⁵ Tunnel vision can depend on how information is presented and structured.¹⁴⁶ For example, in a study where examiners were given a DNA sample to evaluate with case details, they overwhelmingly did not exclude the suspects.¹⁴⁷ Conversely, twelve out of seventeen experts excluded the suspect when reviewing the same sample without receiving any contextual information or case background.¹⁴⁸

Various techniques and practices can mitigate cognitive bias, starting with systematic awareness, understanding high-risk situations, and specific training and operational procedures.¹⁴⁹ Specifically, the complete evaluation and decision-making process should require documentation for use in a trial atmosphere.¹⁵⁰ Mitigation plans should include second-party verification of determined matches, peer review of internal processes, and independent technical review of systems.¹⁵¹ Bias can be mitigated through policies and training, but it must be acknowledged to do so.

142. *Id.* at 539; Rossmo & Pollock, *supra* note 140, at 814.

143. Rossmo & Pollock, *supra* note 140, at 802.

144. *Id.* at 802, 812. *See* Nakhaeizadeh et al., *supra* note 140, at 539 (“The product of various cognitive biases that could obstruct accuracy in what is perceived, how it is perceived, and how it is interpreted is also known in criminal cases as tunnel vision.”).

145. Nakhaeizadeh et al., *supra* note 140, at 537.

146. *Id.* at 542–43 (noting experts evaluate evidence differently when consulting for either defense or prosecution).

147. *Id.* at 545.

148. *Id.*

149. Rossmo & Pollock, *supra* note 140, at 831–32.

150. *Facial Comparison Overview & Methodology Guidelines*, *supra* note 26, at 6.

151. *Id.* at 8.

The concern is not just with individual bias but also with biases in the application of FRT on marginalized communities.¹⁵² Overrepresentation of minorities in operational databases (for example, as a result of the “war on drugs”) can facilitate a negative feedback loop that reinforces racist or biased policing strategies.¹⁵³ For example, the NYPD maintains a database of “gang affiliates” with no requirements for proof or suspicion containing 42,000 images, 99% of which are Black and Latinx individuals.¹⁵⁴ The logic is simple—FRT can only match a photo to another photo in a database, so the more a group is represented, the more likely they are to be a match, especially when the system is less accurate for those same groups. Some critics maintain this is just the most recent installment in a history of racist policing practices.¹⁵⁵

Additionally, overrepresentation and disparate use in the face of known inaccuracies point directly to and perpetuate the enduring system of systemic racism and the over-policing of minority communities.¹⁵⁶ In another example, the Detroit Police Department

152. Najibi, *supra* note 20; see Christopher Jones, Comment, *Law Enforcement Use of Facial Recognition: Bias, Disparate Impacts on People of Color, and the Need for Federal Legislation*, 22 N.C. J.L. & TECH. 777, 788–91 (2021) (discussing how a lack of accurate regulation or transparency surrounding law enforcement’s use of FRT perpetuates biases in policing and disproportionately impacts people of color); Fowler, *supra* note 5 (discussing the built-in racial biases of FRT and how they could be used as a tool to target already marginalized communities); e.g., Aaron Morrison, *50-Year War on Drugs Imprisoned Millions of Black Americans*, AP (July 23, 2021), <https://apnews.com/article/war-on-drugs-75e61c224de3a394235df80de7d70b70> [<https://perma.cc/33L5-TQTR>].

153. See Garvie et al., *supra* note 21.

154. Najibi, *supra* note 20.

155. See Claudia Garcia-Rojas, *The Surveillance of Blackness: From the Trans-Atlantic Slave Trade to Contemporary Surveillance Technologies*, TRUTHOUT (Mar. 3, 2016), <https://truthout.org/articles/the-surveillance-of-blackness-from-the-slave-trade-to-the-police> [<https://perma.cc/2R6Y-D7V4>] (connecting 18th century “lantern laws” in New York City requiring Black and Brown people traveling alone to illuminate their faces as night to be easily identified with the use of artificial light in housing projects).

156. See Drew Desilver, Michael Lipka & Dalia Fahmy, *10 Things We Know About Race and Policing in the U.S.*, PEW RSCH. CTR. (June 3, 2020), <https://www.pewresearch.org/fact-tank/2020/06/03/10-things-we-know-about-race-and-policing-in-the-u-s> [<https://perma.cc/A3Z3-W52Y>]; Damien Patrick Williams, *Fitting the Description: Historical and Sociotechnical Elements of Facial Recognition and Anti-Black Surveillance*, 7 J. RESPONSIBLE INNOVATION S74, S77-80 (2020). See generally SIMONE BROWNE, DARK MATTERS: ON THE SURVEILLANCE OF BLACKNESS (2015). The thoughts of

installed high-definition cameras across the city, allowing them to search video streams against a database containing almost all Michigan residents built from criminal mugshots, driver's licenses, and state IDs.¹⁵⁷ This surveillance heavily correlates with majority-Black areas and is conspicuously absent in majority white and Asian areas.¹⁵⁸ Indeed, many critics have sounded the alarm that disparities will continue to worsen without acknowledgment and specific steps to address them.¹⁵⁹

B. *Admissibility Standards for FRT Evidence*

Facial recognition has all the traditional trappings of eyewitness testimony with a modicum of reliability because it appears akin to forensic evidence.¹⁶⁰ Interestingly, this combination makes a brief discussion of layperson testimony relevant because identification evidence is traditionally layperson. If the facial match is determined to be another form of witness identification, FRT would not be subject to the *Daubert* standard and Federal Rules of Evidence (FRE) Rule 702 but FRE Rule 701. Even so, computer models and forensic evidence are solidly expert; thus, this Section primarily focuses on expert testimony standards.

1. *The distinction between Layperson and Expert*

The distinction between FRE Rule 701 and FRE Rule 702 can be difficult to draw because it is based on the nature of the testimony, not

Black people through their lived experiences should be considered as well. See Emily A. Vogels & Andrew Perrin, *How Black Americans View the Use of Face Recognition Technology by Police*, PEW RSCH. CTR. (July 14, 2022), <https://www.pewresearch.org/fact-tank/2022/07/14/how-black-americans-view-the-use-of-face-recognition-technology-by-police> [<https://perma.cc/G4NC-WQ9F>].

157. Najibi, *supra* note 20; see also REBECCA SMITH, UNIV. MICH. CARCERAL STATE PROJECT, PROJECT GREEN LIGHT: SURVEILLANCE AND THE SPACES OF THE CITY.

158. Najibi, *supra* note 20; see also SMITH, *supra* note 157.

159. E.g., Kade Crockford, *How is Face Recognition Surveillance Technology Racist?*, ACLU (June 16, 2020), <https://www.aclu.org/news/privacy-technology/how-is-face-recognition-surveillance-technology-racist> [<https://perma.cc/26WB-QTLN>]; Smith, *supra* note 67.

160. See Nakhaeizadeh et al., *supra* note 140, at 529, 540 (noting forensic evidence is traditionally regarded as “scientific, objective, and impartial, as well as highly reliable and validated”).

the witness's identity.¹⁶¹ The determination focuses on the facts of the case and the nature of the testimony the witness will offer.¹⁶² Here, a court's decision could turn on how it defines FRT identification—is the testimony provided similar to looking through a mugshot book or reliance on a specialized process, like fingerprint identification?¹⁶³

Generally, courts first look to see if the opinions are on “results from a process of reasoning familiar in everyday life.”¹⁶⁴ Accordingly, courts permit laypersons to provide identification testimony if they are an eyewitness or familiar with the suspect.¹⁶⁵ Some courts allow layperson testimony when the witness's experience or specialized knowledge combines with observations to form a rational opinion.¹⁶⁶ Yet, other courts have narrowed layperson identification testimony by a non-eyewitness to situations where there is no other witness or adequate

161. *United States v. Pembroke*, 876 F.3d 812, 825 (6th Cir. 2017) (permitting lay person testimony about agent's general compilation and use of location data but expert regarding how the data in the reports was captured and analyzed); *see also* *United States v. Natal* 849 F.3d 530, 536 (2d Cir. 2017) (determining testimony about how signals connect to towers required expertise).

162. *Kumho Tire Co. v. Carmichael*, 526 U.S. 137, 148 (1999) (establishing the admissibility test for evidence based on technical expertise).

163. *See* FED. R. EVID. 701 advisory committee's notes to 2000 amendment (FRE Rule 701 is not a means to offer an “expert in lay witness clothing” thus the nature of the testimony is determinative). FRE Rule 701 controls opinion testimony offered by a layperson by limiting the opinion to matters: “(a) rationally based on the witness's perception; (b) helpful to clearly understanding the witness's testimony or to determining a fact in issue; and (c) not based on scientific, technical, or other specialized knowledge within the scope of Rule 702.” FED. R. EVID. 701.

164. *State v. Brown*, 836 S.W.2d 530, 549 (Tenn. 1992); *e.g.*, *United States v. Langford*, 802 F.2d 1176, 1178–79 (9th Cir. 1986) (permitting witnesses with independent knowledge of suspect to provide lay opinions identifying the suspect in surveillance photos).

165. *E.g.*, *United States v. Sanchez*, 789 F.3d 827, 837 (8th Cir. 2015) (concluding extensive surveillance sufficient to support layperson testimony on identity in the blurry video because the expert testimony was likely to help the jury); *United States v. Knowles*, 889 F.3d 1251, 1257 (11th Cir. 2018) (deciding because the suspect was observed during a twenty-minute traffic stop and then identified in surveillance footage by an officer, it was layperson testimony because familiarity with the suspect was sufficient to support FRE 701).

166. *See* *Asplundh Mfg. Div. v. Benton Harbor Engr.*, 57 F.3d 1190, 1201 (3d Cir. 1995); *e.g.*, *United States v. Westbrook*, 896 F.2d 330, 336 (8th Cir. 1990) (accepting chronic amphetamine users as lay witnesses testimony that a substance was amphetamine but excluding another witness with no familiarity from providing the same opinion).

identification.¹⁶⁷ Conversely, when testimony is outside a typical juror's knowledge and involves theories, applications, or conclusions founded in "scientific, technical, or other specialized knowledge," it requires expert qualification.¹⁶⁸ Normally identifying suspects through comparison techniques using surveillance photos or videos requires expert qualification because of the technical or specialized knowledge required.¹⁶⁹

2. *Admissibility standard for expert testimony.*

In *Daubert v. Merrell Dow Pharmaceuticals, Inc.*,¹⁷⁰ the Court determined FRE Rule 702 effectively overruled the "general acceptance" test and offered a non-exhaustive set of factors to determine the admissibility of expert testimony.¹⁷¹ The Court reasoned expert testimony has a highly suggestive nature and significant impact on the jury; thus, judges must be gatekeepers with broad discretion to determine what evidence is admissible based on several factors.¹⁷² The

167. *E.g.*, *United States v. Anderson*, 783 F.3d 727, 747 (8th Cir. 2015). *But see* *United States v. Sellers*, 566 F.2d 884, 886 (4th Cir. 1977) (requiring expert testimony in a situation where there was no other adequate identification testimony).

168. *Kumho Tire Co. v. Carmichael*, 526 U.S. 137, 149 (1999) (applying the *Daubert* standard to technical expert testimony because it similarly is outside knowledge of a layperson like scientific expert testimony).

169. *See, e.g.*, *United States v. Sellers*, 566 F.2d 884, 886 (4th Cir. 1977) (allowing expert testimony on evaluation of photographs including effects of light, shadow, reflections, and distortions where the sole issue in case was the identity of a sole bank robber wearing a disguise); *United States v. Alexander*, 816 F.2d 164, 167–68 (5th Cir. 1987) (allowing expert testimony on identification from photograph to surveillance footage based on nose and mouth area, chin line, ear contours, etc. was proper). *But see* *United States v. Trejo*, 501 F.2d 138, 143 (9th Cir. 1974) (reasoning expert testimony that the suspect was in photographs was not required because it was not beyond the normal observations of the jury).

170. 509 U.S. 579 (1993).

171. *Id.* at 592–95; *see also id.* at 588 (noting nothing in the text of Rule 702 establishes "general acceptance" and there is nothing indicating the Rule intended to incorporate a "general acceptance" standard). The "general acceptance" test was the previous common law standard requiring the party seeking to introduce the expert established theory and method used was generally accepted in the overall field it belonged. *Frye v. United States*, 293 F. 1013, 1014 (D.C. Cir. 1923), *superseded by rule*, *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579 (1993).

172. *Daubert*, 509 U.S. at 592–93 (finding FRE 104(a) controls the testimony's admissibility, but the judge plays a limited, screening role, and the jury is the real decision maker) (citing FED. R. EVID. 104(a)); *see also Kumho Tire*, 526 U.S. at 141 (listing some factors which may be considered, including peer review, potential rate of error, and general acceptance).

specific factors provided by the Court are (1) whether the expert's technique or theory can be or has been tested; (2) whether the technique or theory has been subject to peer review and publication; (3) the scientific technique's known or potential rate of error; (4) the existence and maintenance of standards and controls; and (5) whether the technique or theory has been generally accepted in the scientific community.¹⁷³ Then balancing all factors, judges determine if the underlying theory or technique is "reliable," "scientifically valid," and if the expert's conclusions or methodology "fits" the facts in the case.¹⁷⁴

In making this determination, courts also look to the relevancy and fit requirements of FRE Rule 702 and the balancing test defined by FRE Rule 403.¹⁷⁵ The Court interpreted reliability as the underlying "principles and methods" ability to produce consistent results.¹⁷⁶ Additionally, the validity of the theory, technique, or methodology is just as important as the particular inference that the expert will draw.¹⁷⁷ Subsequently, fit follows closely behind the validity of the theory or technique because the theory can be valid where the application to certain situations is not.¹⁷⁸ Finally, FRE Rule 403 could play an "enhanced role . . . particularly when the scientific or technical

173. *Daubert*, 509 U.S. at 593–94. FRE Rule 702 encapsulates these factors, stating: (a) 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; (b) the testimony is based on sufficient facts or data; (c) the testimony is the product of reliable principles and methods; and (d) the expert has reliably applied the principles and methods to the facts of the case. FED. R. EVID. 702.

174. *Kumho Tire*, 526 U.S. at 141, 150; *Daubert*, 509 U.S. at 589; *see also* Victor E. Schwartz & Cary Silverman, *The Draining of Daubert and the Recidivism of Junk Science in Federal and State Courts*, 35 HOFSTRA L. REV. 217, 221 (2006) (concluding courts must exhibit "a strong and careful judicial gatekeeper function in order to protect a fundamental tenant of justice—finding the truth").

175. FED. R. EVID. 702 (requiring testimony "assist the trier of fact to understand the evidence or to determine a fact at issue"); FED. R. EVID. 403 (balancing "unfair prejudice, confusing the issues, misleading the jury, undue delay, wasting time, or needlessly presenting cumulative evidence" against the probative value); *see also Daubert*, 509 U.S. at 595 (cautioning expert testimony can be "powerful and quite misleading," Rule 403 can exclude otherwise admissible testimony) (citation omitted).

176. *Daubert*, 509 U.S. at 590 n.9.; *see also* FED. R. EVID. 702(c); Edward J. Imwinkelried, *The Best Insurance Against Miscarriages of Justice Caused by Junk Science: An Admissibility Test that is Scientifically and Legally Sound*, 81 ALB. L. REV. 851, 864–65 (2017).

177. *See* Gen. Elec. Co. v. Joiner, 522 U.S. 136, 144 (1997).

178. *See id.*

knowledge proffered is novel or controversial.”¹⁷⁹ Ultimately, the judge as the gatekeeper balances all of the factors and applies FRE Rule 702 and 403 to make a final determination to admit or bar testimony.¹⁸⁰

The Section below will detail how courts determine admissibility by discussing each *Daubert* factor, FRE Rule 702, and FRE Rule 403.

a. Testing of the theory or technique

In evaluating a theory or technique, the thrust of the inquiry is whether it *can be* tested and *has it* been tested.¹⁸¹ Initially, courts evaluate if the theory or technique can be tested because “while hypotheses and theories may be freely invented and *proposed* in science, they can be *accepted* . . . only if they pass critical scrutiny, which includes . . . careful observation or experiment.”¹⁸² Hypotheses that have not been tested (i.e. validated) through experiments are less secure and tend toward inadmissibility.¹⁸³ In some cases, once a court determines a theory is testable and testing has occurred, that is the end of the evaluation.¹⁸⁴ For example, in *United States v. Hunt*¹⁸⁵ and *United*

179. *United States v. Posado*, 57 F.3d 428, 435 (5th Cir. 1995) (evaluating if polygraph evidence satisfies Rule 702, which the court did not determine, other evidentiary rules could still exclude the evidence).

180. *Joiner*, 522 U.S. at 146; *Kumho Tire Co. v. Carmichael*, 526 U.S. 137, 141 (1999).

181. *Daubert*, 509 U.S. at 593–94 (finding that “the criterion of the scientific status of a theory is its . . . refutability”) (quoting K. Popper, *Conjectures and Refutations: The Growth of Scientific Knowledge* 37 (5th ed. 1989)).

182. CARL G. HEMPEL, *PHILOSOPHY OF NATURAL SCIENCE* 16 (1966).

183. See *United States v. Gissantaner*, 990 F.3d 457, 463–64 (6th Cir. 2021) (“An untestable scientific theory is all theory and no science. In the absence of proof that a technology ‘can be . . . tested,’ there is no way to show whether it works . . . and no way to give it ‘scientific status.’”) (citations omitted); see, e.g., *Smith v. Cangieter*, 462 F.3d 920, 924 (8th Cir. 2006) (testifying mechanical expert using a novel theory that under certain road conditions a four-wheel drive system experiences mechanical failures did not have any accident statistics or mathematical models to support his conclusion in automobile products liability action). Some courts deem testability wholly satisfied after simply determining a theory can be tested. *Lee v. Martinez*, 96 P.3d 291, 299 (N.M. 2004) (satisfying *Daubert* prong simply because “the control question polygraph examination can be tested”); see *United States v. Mitchell*, 365 F.3d 215, 238 (3d Cir. 2004) (finding the fact that fingerprint identifications are testable as enough to weigh in favor of admissibility).

184. E.g., *United States v. Bonds*, 12 F.3d 540, 558 (6th Cir. 1993) (concluding “the FBI’s principles and methodology have in fact been tested” by proficiency testing and validation studies).

185. 442 F. Supp. 3d. 762 (E.D.N.Y. 2019).

*States v. Shipp*¹⁸⁶ the courts examined the theory and science supporting firearm tool identification.¹⁸⁷ The *Hunt* court concluded that testability was satisfied because the theory can be and has been tested.¹⁸⁸ The *Shipp* court concluded the same and reasoned further, “the probative value of different study designs is more appropriately considered as part of the discussions of the method’s error rate.”¹⁸⁹ Ultimately, testability aims to distinguish objective theory from subjective judgments, not parse test results or methods.¹⁹⁰

Some courts employ a more nuanced approach, reasoning valid testing requires verification that testing produces consistent results.¹⁹¹ For example, in *United States v. Morgan*,¹⁹² the court determined the admissibility of expert testimony regarding a mixed DNA sample.¹⁹³ The court reasoned there is no requirement to conduct the studies in all possible scenarios, and extensive testing is not required if a sound conclusion can be made.¹⁹⁴ However, in *United States v. Mitchell*,¹⁹⁵ the court cautioned that the FBI experiment on fingerprints did not accurately simulate distorted, real world prints concluding it wasn’t strong support for testability.¹⁹⁶ The difference between *Morgan* and *Mitchell* rests not in the test results but in the “good grounds” for the result. Accordingly, when evaluating testing, the proper focus is if the testing supports validity, not if the results are accurate.¹⁹⁷

186. 502 F. Supp. 3d. 28 (D.C. Cir. 2020).

187. *Id.* at 37; *Shipp*, 442 F. Supp. 3d. at 776; *see e.g.* *United States v. Harris*, 464 F. Supp. 3d. 1252 (W.D. Okla. 2020).

188. *Hunt*, 464 F. Supp. 3d. at 1257.

189. *Shipp*, 442 F. Supp. 3d. at 776.

190. *Id.*; *Harris*, 464 F. Supp. 3d.

191. *Harris*, 464 F. Supp. 3d. at 37; *United States v. Taylor*, 663 F. Supp. 2d 1170, 1176 (D.N.M. 2009).

192. 53 F. Supp. 3d 732 (S.D.N.Y. 2014), *aff’d*, 675 F. App’x 53 (2d Cir. 2017).

193. *Id.* at 734.

194. *Id.* at 744; *see United States v. Mitchell*, 365 F.3d 215, 237 n.18 (3d Cir. 2004) (reasoning to extrapolate the experiment out to real-world conditions such as the FBI’s full database would take millennia to complete).

195. 365 F.3d 215 (3d Cir. 2004).

196. *Id.* at 238.

197. *See Kumho Tire Co. v. Carmichael*, 526 U.S. 137, 153 (1999) (reasoning if the conclusion is within “the range where experts might reasonably differ,” it goes to the weight, not admissibility); *see also Guild v. General Motors Corp.*, 53 F. Supp. 2d 363, 369 (W.D.N.Y. 1999) (“Such differences of opinion and alleged weaknesses in the experts’ methodologies will go to the weight to be given the expert testimony, not its admissibility.”).

b. Peer review and publication

Peer reviews ensure the vetting of emerging theories and techniques for errors, but it is not essential to admissibility.¹⁹⁸ Until theories are public, there is no way for the relevant scientific community to probe the theory; thus, a court should regard it with skepticism.¹⁹⁹ A key factor of this analysis is the type of publication, which should be well respected and conduct a review of methodology as part of the publication process.²⁰⁰ Some courts have conflated the independent review of a second examiner as the peer review referred to in *Daubert*. For example, in *Mitchell*, the court balanced a second examiner review with the contention that groupthink rendered the evaluations unreliable, concluding it “may not be peer review in its best form, but, on balance [it] . . . does favor admission.”²⁰¹ However, if the procedure itself is not validated, this produces two wrong results, not the discriminating evaluation of a theory contemplated by *Daubert*.²⁰²

198. The Court was careful to distinguish that peer review is not a requirement as some theories are well grounded without it, and some are too novel. *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579, 593 (1993). Nevertheless, peer review is important because it increases the likelihood of discovering substantial flaws in the methodology. *Id.*

199. Francisco J. Ayala & Bert Black, *Science and the Courts*, 81 AM. SCIENTIST 230, 238 (1993).

200. *United States v. Prime*, 363 F.3d 1028, 1034 (9th Cir. 2004) (citing “numerous journals . . . peer review[ed] by not only handwriting experts, but others in the forensic science community”); see *United States v. Adams*, 444 F. Supp. 3d 1248, 1265–66 (D. Or. 2020) (concluding the purpose for peer review in *Daubert* was not met because the Association of Firearm and Tool Mark Examiners Journal is a “trade publication, meant only for industry insiders, not the scientific community. Second . . . the purpose of publication . . . is not to review the methodology for flaws”); accord *United States v. Tibbs* No. 2016-CF1-19431, 2019 WL 4359486, at *25 (D.C. Super. Ct. Sept. 5, 2019). But see *United States v. Hunt*, 464 F. Supp. 3d 1252, 1257 (W.D. Okla. 2020) (relying on articles in the Association of Firearm and Tool Mark Examiners Journal as peer review).

201. *United States v. Mitchell*, 365 F.3d 215, 239 (3d Cir. 2004).

202. *United States v. Llera Plaza*, 179 F. Supp. 2d 492, 509 (E.D. Pa.) (stating “another opinion rendered by another examiner, whether in corroboration or in refutation, does little to put a ‘scientific’ gloss on the first opinion, much less constitute ‘peer review’”), *vacated*, 188 F. Supp. 2d 549 (E.D. Pa. 2002); accord *United States v. Sullivan*, 246 F. Supp. 2d 700, 704 (E.D. Ky. 2003). But see *Mitchell*, 365 F.3d at 238 (distinguishing “verification step” from research but concluding it is akin to peer review because in this case, it can be a retesting of the process, not simply confirming a match)).

c. *Rates of error & controlling standards*

The *Daubert* Court included both the “known or potential rate of error” and “the existence of maintenance of standards controlling the technique’s operation” in the same paragraph; thus, courts typically evaluate them in tandem.²⁰³ The two are interrelated because the absence of clear standards increases subjectivity, which can lower accuracy.²⁰⁴ The scientific community has long recognized that standards and controls are crucial to create reliable, repeatable results because it ensures the estimation of an error rate applies consistently across examiners or systems.²⁰⁵ In this way, strict standards and controls can mitigate error rates.

Accreditation and licensing, which usually require compliance with procedures and auditing or oversight, are convincing facts supporting admissibility.²⁰⁶ For example, in *United States v. Beasley*,²⁰⁷ the court reasoned the DNA lab’s accreditation, protocols that conformed to guidelines accepted by the scientific community, and double verification supported accurate results.²⁰⁸ Despite Beasley’s allegations of lack of proficiency testing, double blind tests, and maintaining records of errors, the court determined these facts went to the weight not admissibility of the expert testimony.²⁰⁹ However, the court cautioned when reliable methods are altered to the point of “skew[ing] the methodology itself,” they can be excluded.²¹⁰

Critical considerations for assessing accuracy are the quantity and quality of the data, and the “known or potential rate of error,”

203. *Daubert*, 509 U.S. at 594; see Imwinkelried, *supra* note 176, at 865 (reiterating if the testimony is based on a scientific foundation, there must be a showing of accuracy).

204. *Pettus v. United States*, 37 A.3d 213, 224–45 (D.C. 2012) (relying on the fact that FBI document examiners are trained and employ national standards by ASTM International); see *United States v. Shipp*, 422 F. Supp. 3d 762, 779–82 (E.D.N.Y. 2019) (determining the “near total subjectivity” of firearms-toolmark identification does not meet controlling standards and weighs against admissibility).

205. *Daubert*, 509 U.S. at 593; e.g., *United States v. Crisp*, 324 F.3d 261, 269 (4th Cir. 2003); see *United States v. Williams*, 583 F.2d 1194, 1198 (2d Cir. 1978) (ruling ten matching features must be found in voice spectra prior to positive identification).

206. *United States v. Beverly*, 369 F.3d 516, 530 (6th Cir. 2004) (noting labs undertaking DNA forensic work are accredited by the American Society of Crime Laboratory directors, an external agency).

207. 102 F.3d 1440 (8th Cir. 1996).

208. *Id.* at 1447.

209. *Id.* at 1448.

210. *Id.*

specifically false negatives and positives.²¹¹ The Third Circuit in *Mitchell* acknowledged these two error rates before reasoning the false positive rate was more important because it translated to the ability to make a positive identification.²¹² Additionally, the data quantity and quality help establish accurate test conditions reducing the gap between laboratory tests and actual conditions.²¹³

Notably, the *Daubert* error rate analysis can shift from a focus on standards to the expert's experience as the underlying discipline of the expert's testimony relies less on scientific and more on specialized experience, like latent fingerprint analysis.²¹⁴ For example, in *United States v. Havvard*,²¹⁵ a fingerprint expert testified there was a small margin of error and no single standard for determining if there were enough points for comparison.²¹⁶ But, the court cited the expert's experience and the peer review process of double checking the results as support for admissibility because both minimize error.²¹⁷ Similarly, handwriting analysis can be admissible based on the witness's extensive experience and scholarship in addition to standardized evaluation methods and techniques.²¹⁸ However, in these cases, judges should not accept an expert's bare assertion that the theory or technique is valid because of the expert's experience.²¹⁹ For example, in *United States v. Saelee*,²²⁰ the court focused on the science of handwriting analysis, not

211. *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579, 594 (1993); see FED. R. EVID. 702(b) (asking if testimony is based on "sufficient facts or data?"). See generally Imwinkelried, *supra* note 176, at 862–63.

212. 365 F.3d 215, 239 (3d Cir. 2004).

213. *United States v. Bonds*, 12 F.3d 540, 560 (6th Cir. 1993) (noting the "deficiencies in calculating the rate of error and the failure to conduct external blind proficiency tests are troubling").

214. *Kumho Tire Co. v. Carmichael*, 526 U.S. 137, 141 (1999); see also Edward J. Imwinkelried, *The Next Step After Daubert: Developing a Similarly Epistemological Approach to Ensuring the Reliability of Nonscientific Expert Testimony*, 15 CARDOZO L. REV. 2271, 2289 (1994) (supporting the premise that experience is the basis of expertise).

215. 260 F.3d 597 (7th Cir. 2001)

216. *Id.* at 599.

217. *Id.* at 598–99. Also important to the court's determination was that fingerprint analysis has been used for "roughly 100 years". *Id.* at 598.

218. *United States v. Jones*, 107 F.3d 1147, 1159–60 (6th Cir. 1997).

219. *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579, 590 (1993); *Gen. Elec. Co. v. Joiner*, 522 U.S. 136, 144–45 (1997); see, e.g., *Jones*, 107 F.3d at 1159 (citing that the Third Circuit, Southern District of New York, and DC Circuit permit this testimony).

220. 162 F. Supp. 2d. 1097 (D. Alaska 2001).

the expert's experience, although it was a factor.²²¹ After reviewing the expert's description of the process, the court determined the technique "appears to be entirely subjective and entirely lacking in controlling standards."²²² The court reasoned that without more than the expert's description, the testimony "would itself be nothing more than a set of subjective observations."²²³

d. General acceptance

The evaluation of widespread acceptance harkens back to *Frye v. United States*²²⁴ and usually determines the weight of the evidence rather than admissibility.²²⁵ General acceptance is just that, general; it does not require absolute certainty.²²⁶ The existence of a "substantial portion of the pertinent scientific community" which supports or accepts the underlying science of the "theory, principles, and methodology" is sufficient.²²⁷ If a theory remains in the "twilight zone" due to novelty or newness, other factors can dominate, but the admissibility of a discredited theory is nearly impossible.²²⁸ One need look no further than *Frye* and the history of lie detector test admissibility to illustrate the near impossibility.²²⁹ It took seventy years for the court to overturn the per se rule of inadmissibility in *United States v. Posado*,²³⁰ and simply determine the *Daubert* decision demanded a full evaluation of polygraph testing theory.²³¹ The court required a *Daubert* evaluation of polygraph evidence because much had changed, such as increased biological measures and

221. *Id.* at 1104.

222. *Id.*

223. *Id.* at 1105.

224. 293 F. 1013 (D.C. Cir. 1923), *superseded by rule*, *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579 (1993).

225. *Daubert*, 509 U.S. at 594.

226. *United States v. Bonds*, 12 F.3d 540, 561 (1993).

227. *Id.*

228. *Frye*, 293 F. at 1014; *see also Bonds*, 12 F.3d at 562 (reasoning "'neither consensus nor certainty' is needed, an absence of consensus is not immaterial") (citation omitted). For an additional example of a "twilight" case post-*Daubert*, *see Lee v. Martinez*, 96 P.3d 291 (N.M. 2004).

229. *Frye* arose from a murder defendant seeking to use a predecessor of the polygraph to prove innocence. *See* 293 F. at 1013.

230. 57 F.3d 428 (5th Cir. 1995).

231. *Id.* at 436.

professionalization of the field.²³² But, the court cautioned, “[i]t is with a high degree of caution that we have today opened the door . . . [w]e may indeed be opening a legal Pandora’s box.”²³³

Similarly, courts have historically rejected eyewitness experts because research is incomplete and insufficient to determine validity.²³⁴ Over time, some courts have noted the increase of modern scientific studies showing the questionable nature of eyewitness testimony based on widely accepted methodology.²³⁵ These courts have concluded that expert testimony on the subject can be admissible because the science of eyewitness perception has achieved the level of “exactness, methodology and reliability of any psychological research.”²³⁶

e. FRE Rule 702: relevancy & fit

FRE Rule 702 is a two-part consideration where the expert’s relevant conclusion is considered in light of the facts presented, not a generalized understanding of the field.²³⁷ The first part of the inquiry, commonly referred to as the relevancy requirement, asks, will it help the jury determine a fact at issue?²³⁸ Courts distinguish between evidence the jury can adequately weigh and evidence that is outside common knowledge, often admitting the latter and excluding the

232. *Id.* at 434.

233. *Id.* at 436.

234. *E.g.*, *United States v. Kime*, 99 F.3d 870, 883–84 (8th Cir. 1996).

235. *E.g.*, *United States v. Smithers*, 212 F.3d 306, 311 (6th Cir. 2000); *Minor v. United States*, 57 A.3d 406, 420 (D.C. 2012); *cf.* Saul M. Kassin, V. Anne Tubb, Harmon M. Hosch & Amina Memon, *On the “General Acceptance” of Eyewitness Testimony Research*, 56 AM. PSYCHOLOGIST 405, 405 (2001) (surveying sixty-four eyewitness experts). *See generally* Roger B. Handberg, *Expert Testimony on Eyewitness Identification: A New Pair of Glasses for the Jury*, 32 AM. CRIM. L. REV. 1013, 1018–22 (1995).

236. *See United States v. Smith*, 736 F.2d 1103, 1106, 1108 (6th Cir. 1984) (per curiam); *Smithers*, 212 F.3d at 312, 315–16.

237. *Gen. Elec. Co. v. Joiner*, 522 U.S. 136, 146 (1997); *see, e.g.*, *Newman v. Motorola, Inc.*, 218 F. Supp. 2d 769, 783 (D. Md. 2002) (rejecting expert research on cell phones increasing the risk for brain cancer because plaintiff had the wrong subtype of brain tumor); *In re Paoli R.R. Yard PCB Litig.*, 35 F.3d 717, 744 (3d Cir. 1994) (“[F]or animal studies to be admissible to prove causation in humans, there must be good grounds to extrapolate from animals to humans.”).

238. *Joiner*, 522 U.S. at 141.

former.²³⁹ For example, courts resoundingly determine DNA evidence placing a suspect at the scene of the crime in question is relevant.²⁴⁰ Conversely, courts have excluded defense eyewitness expert testimony, reasoning it allows an expert to “comment on the weight and credibility of opponents’ witnesses.”²⁴¹ However, it can be admissible when expert testimony is carefully tailored to a discussion on the science of eyewitness identification and not on the specific eyewitness.²⁴² Here, the distinction is that the underlying psychology supporting decreased reliability of eyewitness identifications is counterintuitive to juries; thus, it is helpful and relevant.²⁴³

The second part of the fit inquiry asks, is the conclusion or purpose of the expert’s testimony applicable to the facts of the case?²⁴⁴ This is inherent in FRE Rule 702(d), which requires a judge to determine if the conditions during experiments or field observations accurately approximate the facts of a case.²⁴⁵ Courts have cautioned that just because one application and conclusion is accepted and reliable does not mean extrapolating or applying it to another is equally so.²⁴⁶ The determination is not based on the correctness of an expert’s conclusion, only if it is reasonable, which ensures the jury’s province is

239. *See, e.g.*, *United States v. Beasley*, 102 F.3d 1440, 1447 (8th Cir. 1996) (agreeing with the district court that the DNA evidence addressed “the factual issue of whether” the defendant was the person in the mask that committed a bank robbery).

240. *E.g., id.*

241. *United States v. Smith*, 122 F.3d 1355 (11th Cir. 1997) (citation omitted); *see also United States v. Brien*, 59 F.3d 274, 277 (1st Cir. 1995); *United States v. Harris*, 995 F.2d 532, 534 (4th Cir. 1993) (“Until fairly recently, most, if not all, courts excluded expert psychological testimony on the validity of eyewitness identification.”).

242. *Harris*, 995 F.2d at 535; *United States v. Stevens*, 935 F.2d 1380, 1400 (3d Cir. 1991).

243. *Harris*, 995 F.2d at 535; *Stevens*, 935 F.2d at 1400.

244. *See Gen. Elec. Co. v. Joiner*, 522 U.S. 136, 144–45 (1997) (using this framework to determine if the expert’s testimony was related to the facts of the case).

245. FED. R. EVID. 702(d); *see also Imwinkelried, supra* note 176, at 863 (“[T]he trial judge must decide whether, as a matter of logic, the expert may extrapolate from the empirical studies to an inference about the case.”) (citing *Joiner*, 522 U.S. at 146).

246. *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579, 591 (1993); *see also Joiner*, 522 U.S. at 146 (observing “[a] court may conclude that there is simply too great an analytical gap between the data and the opinion proffered”).

not invaded.²⁴⁷ Judges are gatekeepers, but juries are the ultimate determiner of fact.²⁴⁸

Finally, an otherwise reliable scientific principle can be undermined “if ‘a reliable methodology was so altered . . . as to skew the methodology itself’” or if the analytical gap from theory to facts is simply too great.²⁴⁹ For example, when evaluating the PCR technique for DNA testing, the court in *Beasley* reasoned lack of proficiency testing, double-blind external tests, and maintaining records of errors did not “so alter[.]” the DNA PCR methodology making it inadmissible.²⁵⁰ Instead, these concerns went to the weight of the evidence. However, the court did elaborate further, indicating challenges can occur “by showing . . . sloppy handling of the samples, failure to properly train those performing testing, failure to follow appropriate protocols, and the like.”²⁵¹

3. *Determining admissibility: balancing the Daubert factors*

A court’s determination on admissibility using the *Daubert* analysis often concludes by balancing all the factors against each other. *Daubert* posits that expert testimony must be based on solid footing to allow the adversarial process to address inadequacies through competing expert testimony and vigorous cross-examination.²⁵² The *Mitchell* court succinctly captured this sentiment stating,

Indeed, as our discussion of the various *Daubert* factors suggests, many of them are guarantees that cross-examination and adversary testing will be possible: Testability ensures the basic possibility of meaningful cross-examination. Peer review and publication also provide raw material for the cross-examining attorney to confront the expert with. The existence of a known error rate may force an expert to admit to the limitations of his or her methods. The maintenance of standards provides

247. FED. R. EVID. 702 advisory committee’s notes to 2000 amendment; Imwinkelried, *supra* note 176, at 867.

248. FED. R. EVID. 702 advisory committee’s notes to 2000 amendment.

249. *Beasley*, 102 F.3d at 1448 (citation omitted).

250. *Id.*

251. *Id.*

252. *Daubert*, 509 U.S. at 596 (“Vigorous cross-examination, presentation of contrary evidence, and careful instruction on the burden of proof are the traditional and appropriate means of attacking shaky by admissible evidence.”); accord *United States v. Mitchell*, 365 F.3d 215, 244 (3rd Cir. 2004).

an objective benchmark to confirm that the expert did indeed follow her method.²⁵³

The *Mitchell* court concluded that latent fingerprint analysis evidence was admissible despite a lack of objective standards because low estimations of error balanced with a peer review process supported reliability.²⁵⁴ Further, testability and general acceptance also helped admissibility providing a firm basis for cross-examination.²⁵⁵

In another example, the *Morgan* court relied heavily on the stability of the underlying scientific field of DNA evaluation supported by a showing of a relevant peer review process.²⁵⁶ The court determined the testimony was admissible even though no other public lab, including the FBI, found it reliable enough to use in criminal cases because of shortcomings such as internal threshold numbers and an understanding of the false positives based on race.²⁵⁷ The court reasoned the software itself had been tested and was based on sound mathematical principles.²⁵⁸ Despite the incomplete evaluation, *Daubert* does not require complete testing of all scenarios because there was “sufficient scientific underpinning” to allow a jury to determine the weight of the evidence.²⁵⁹

Finally, the example of firearm toolmark analysis is instructive because it has not passed *Daubert* in some cases.²⁶⁰ Generally, courts accept the method is testable and is published in a variety of scientific and academic sources.²⁶¹ Even though the standards are highly subjective because of a “relatively low rate of error,” admissibility is warranted.²⁶² This can be and often is further supported by actual evidence that, in the particular case, the methodology was applied

253. *United States v. Mitchell*, 365 F.3d. 215, 245 (3rd Cir. 2004).

254. *Id.* at 241.

255. *Id.* at 239, 241.

256. 53 F. Supp. 3d 732, 741–47 (S.D.N.Y. 2014), *aff'd*, 675 F. App'x 53 (2d Cir. 2017) (reasoning the state lab was independently accredited, presented numerous scientific articles and presentations at scientific gatherings, and despite dissenting voices, there is a sufficient scientific underpinning).

257. *Id.* at 742–43.

258. *Id.*

259. *Id.* at 743 (contending this is the type of determination a jury should make).

260. *United States v. Harris*, 502 F. Supp. 3d 28, 43 (D.C. Cir. 2020); *United States v. Shipp*, 422 F. Supp. 3d 762, 783 (E.D.N.Y. 2019).

261. *Harris*, 502 F. Supp. 3d at 43.

262. *Id.*

correctly by an experienced examiner.²⁶³ Conversely, some courts have come down on the other side, citing the lack of a precise error rate, subjective standards, and some disagreement over the validity of the underlying theory.²⁶⁴ On balance, these factors led the court to conclude the method is not reliable enough to determine “to any degree of certainty that the recovered firearm *is* the source of the recovered bullet fragment.”²⁶⁵ Ultimately, where the balancing test is close, the facts of the case can tip the scale.

II. ANALYSIS

FRT identification evidence should require expert testimony because the basis of the witness’s conclusion is beyond the knowledge of a typical juror.²⁶⁶ First, like video surveillance identification which requires expert testimony, a witness using FRT would rely on experience and training in facial comparison techniques.²⁶⁷ Narrowly, one could argue it is layperson testimony because anyone can compare two photos and determine if it is the same person. But courts have been clear—absent familiarity with the suspect; the conclusion relies on specialized knowledge and training in identifying suspects.²⁶⁸ Second, FRT is based on specialized knowledge of a scientific theory which

263. *Id.*

264. *Shipp*, 422 F. Supp. 3d at 783.

265. *Id.*

266. *Kumho Tire Co. v. Carmichael*, 526 U.S. 137, 149 (1999) (applying the *Daubert* standard to technical expert testimony because it similarly is outside knowledge of a layperson like scientific expert testimony).

267. See McCoy, *supra* note 25, at 488–89; John Nawara, Note, *Machine Learning: Face Recognition Technology Evidence in Criminal Trials*, 49 U. LOUISVILLE L. REV. 601, 614, 620–21 (2011). Compare *United States v. Sellers*, 566 F.2d 884, 886 (4th Cir. 1977) (allowing expert testimony on evaluation of photographs) and *United States v. Alexander*, 816 F.2d 164, 167–68 (5th Cir. 1987) (allowing expert testimony on identification from photograph to surveillance footage based on nose and mouth area, chine line, ear contours, etc. was proper) with *United States v. Langford*, 802 F.2d 1176, 1178–79 (9th Cir. 1986) (permitting witnesses with independent knowledge of suspect to provide lay opinions identifying the suspect in surveillance photos).

268. *New York v. Reyes*, 133 N.Y.S. 3d 433, 434–36 (N.Y. Sup. Ct. 2020) (discussing a law enforcement officer’s review of results from facial recognition with other materials, then view surveillance footage could be admissible without expert qualification); see McCoy, *supra* note 25, at 488–89 (reasoning “similar to a police officer standing in a crowd with a stack of mug shots and comparing them to people who walk past him” the technology simply enhances the human skill of identifying, making the process more efficient).

further indicates it should be subject to expert testimony.²⁶⁹ Correspondingly, FRT also resembles DNA evidence which uses scientific principles and software to determine matches that an expert then evaluates.²⁷⁰ Testimony for DNA identifications is solidly expert because the underlying scientific theory is beyond a typical juror's understanding.²⁷¹ Similarly, the scientific theory of facial analysis and comparison is not common knowledge, meaning FRT identifications should also require expert testimony. Consequently, this Section proceeds on the premise that expert testimony is necessary to use FRT results as proof of identity in a criminal trial.

The differences between the two historical events of January 6th and the BLM protests frame the analysis in this Section. Specifically, the differing racial composition of the participants, the probe photograph quality, and the technical sophistication of the investigating authority and software are all critical factors of the analysis. First, many of the January 6th participants were white, middle-aged men, which matches the race and gender of people most accurately identified by FRT.²⁷² Additionally, the probe images were taken during daylight hours or inside well-lit areas of the United States Capitol.²⁷³ The images captured on high-quality surveillance video often produced multiple

269. See McCoy, *supra* note 25, at 621.

270. Cf. *United States v. Havvard*, 260 F.3d 597, 598 (7th Cir. 2001) (admitting expert testimony about the process of lifting fingerprints from a handgun that were later to the defendant); *United States v. Pembroke*, 876 F.3d 812, 825–26 (6th Cir. 2017) (determining lay witness testimony was permitted for cell tower location data related to the agent's use of the reports, and the general compilation of the reports contrasting with the expert testimony was related to how cell-site data is captured and analyzed); see also *United States v. Natal*, 849 F.3d 530, 536 (2d Cir. 2017) (per curiam) (joining the Seventh and Tenth Circuits that expert testimony is required to testify about how cell phone towers operate).

271. *McDaniel v. Brown*, 588 U.S. 120, 127 (2010); accord *United States v. Bonds*, 12 F.3d 540, 546 (6th Cir. 1993), *United States v. Martinez*, 3 F.3d 1191, 1192 (8th Cir. 1993), *United States v. Brooks*, 678 Fed. Appx. 755 (10th Cir. 2017).

272. Scott Tong & Serena McMahon, *White, Employed and Mainstream: What We Know About the Jan. 6 Rioters One Year Later*, WBUR (Jan. 3, 2022), <https://www.wbur.org/hereandnow/2022/01/03/jan-6-rioters-white-older> [<https://perma.cc/4NQM-VABX>].

273. See N.Y. Times, *Day of Rage: How Trump Supporters Took the U.S. Capitol | Visual Investigations*, YOUTUBE (Jul. 1, 2021), <https://www.youtube.com/watch?v=jWJVMoe7OY0> [<https://perma.cc/6PAY-6CKQ>] (compiling the numerous videos of the attack on the capitol grounds).

images from different angles in close quarters.²⁷⁴ Finally, federal investigators benefiting from practical experience, specialized training, and FBI technological resources analyzed the images using clear standards and procedures.²⁷⁵

Conversely, the BLM protests had none of these factors at play. For example, the BLM protests in New York, Miami, and Philadelphia included a large number of Black and Brown protestors.²⁷⁶ While individuals of all races participated in the demonstrations, up to 46% identified as Black or Brown—the demographic group least likely to generate an accurate FRT match.²⁷⁷ Next, the probe photographs of the protestors came from cell phones or CCTV cameras, which produce degraded image quality compared to high-resolution surveillance cameras.²⁷⁸ The quality of the photographs was further undermined because the images were taken outdoors and sometimes at night.²⁷⁹ Finally, the state and local agencies that analyzed the images running them through FRT software were less experienced and less resourced than their federal counterparts.²⁸⁰ Each of these factual

274. *Id.*

275. *Infra* Section II.C.2.

276. Dana R. Fisher, *The Diversity of the Recent Black Lives Matter Protests is a Good Sign for Racial Equity*, BROOKINGS: HOW WE RISE (July 8, 2020), <https://www.brookings.edu/blog/how-we-rise/2020/07/08/the-diversity-of-the-recent-black-lives-matter-protests-is-a-good-sign-for-racial-equity> [<https://perma.cc/X9CF-ZW8R>].

277. *News Report Reveals Demographics of Black Lives Matter Protesters Shows Vast Majority Are White, Marched Within Their Own Cities*, PR NEWswire (June 18, 2022, 8:38 AM), <https://www.prnewswire.com/news-releases/new-report-reveals-demographics-of-black-lives-matter-protesters-shows-vast-majority-are-white-marched-within-their-own-cities-301079234.html> [<https://perma.cc/9V44-56LR>].

278. Jake Seiner, *Black Lives Matter Movement: A Case for Camera Phones?*, CHRISTIAN SCI. MONITOR (June 17, 2020), <https://www.csmonitor.com/USA/Society/2020/0617/Black-Lives-Matter-movement-A-case-for-camera-phones> [<https://perma.cc/K25M-6VBR>] (noting that one lawyer publishing clips from BLM protests found 80% of his early posts were from non-journalists, often from cell phones); Press Release, ACLU, *Ban Dangerous Facial Recognition Technology That Amplifies Racist Policing* (Jan. 26, 2021), <https://www.amnesty.org/en/latest/press-release/2021/01/ban-dangerous-facial-recognition-technology-that-amplifies-racist-policing> [<https://perma.cc/BK2R-6F2K>].

279. Seiner *supra* 278, Taylor *supra* 3.

280. Russel Brandom, *Facebook, Twitter, and Instagram Surveillance Tool Was Used to Arrest Baltimore Protestors*, THE VERGE (Oct. 11, 2016, 1:42 PM), <https://www.theverge.com/2016/10/11/13243890/facebook-twitter-instagram>

differences will be addressed in Sections II.A–E through a *Daubert* analysis of these two scenarios.

A. *Testing of the Theory or Technique*

If the only question is, can FRT be tested—the answer is yes because the NIST conducted repeated tests using multiple FRT software programs.²⁸¹ This fact alone can support admissibility when the theory has a history of admissibility and acceptance in the courtroom, which is not the case for FRT.²⁸² Consequently, a court should move beyond the simple “test or no test” inquiry to evaluate FRT testability by examining the ability of testing to produce consistent and reliable results.²⁸³ This evaluation ensures that FRT testimony is not a “subjective conclusion eluding assessment.”²⁸⁴ Turning to apply this standard to the NIST testing—consistent and reliable results are why it is considered the industry standard for FRT testing.²⁸⁵ In fact, the NIST tested the FBI’s and NYPD’s Clearview AI software in a controlled black box setting, referred to as the gold standard for producing objective and reliable results.²⁸⁶ The NIST black box testing has been repeated over a decade, confirming the consistent and reliable test methods of

police-surveillance-geofeedia-api [https://perma.cc/XUF3-QN9B]; Tawana Petty, *Defending Black Lives Means Banning Facial Recognition*, WIRED, (Jul. 10, 2020, 8:00 AM), https://www.wired.com/story/defending-black-lives-means-banning-facial-recognition [https://perma.cc/852Q-PUQM].

281. See *supra* notes 100–107 and accompanying text.

282. *United States v. Hunt*, 464 F. Supp. 3d 1252, 1256–57 (W.D. Okla. 2020).

283. *United States v. Harris*, 502 F. Supp. 3d 28, 37 (D.C. Cir. 2020).

284. FED. R. EVID. 702, advisory committee note to 2000 amendment; *United States v. Harris*, 502 F. Supp. 3d 28, 37 (D.D.C. 2020); *United States v. Mitchell*, 365 F.3d 215, 245 (3rd Cir. 2004).

285. Press Release, Idemia, IDEMIA’s Facial Recognition Ranked #1 in NIST’s Latest FRVT Test (Apr. 6, 2021) https://www.idemia.com/press-release/idemias-facial-recognition-ranked-1-nists-latest-frvt-test-2021-04-06 (calling FRVT test results the “gold standard of the global security industry”); see e.g., INNOVATRICES, *Innovatrics: Top Performer in Every NIST FRVT Category*, https://www.innovatrics.com/awards/frvt/ (last visited Dec. 4, 2022); NEC, *NIST Validation Third-Party Authenticated*, https://www.necam.com/AdvancedRecognitionSystems/NISTValidation/Fingerprin tFacial/ (last visited Dec. 4, 2022).

286. *Hearings*, *supra* note 50, at 5 (statement of Kimberly J. Del Greco, Deputy Assistant Director, FBI); *Harris*, 502 F. Supp. 3d at 38 (supporting testability prong of *Daubert* for firearm toolmark identification with multiple black box tests that show the method is valid and reliable).

FRT algorithms and methodologies.²⁸⁷ This heightened evaluation verifies that NIST tests form a basis for effective cross-examination relegating questions of test sufficiency to the adversarial process.²⁸⁸

Testability of all FRT algorithms is not satisfied by the NIST testing for two reasons, the lack of industry standards and source code. First, the lack of industry standards means FRT algorithms can use discredited, unreliable methods.²⁸⁹ Because the underlying methodology of FRT algorithms can vary, there is no “careful observation” and “critical scrutiny” without testing.²⁹⁰ Second, most source code is not open-source, making it challenging, if not impossible, to draw an analogous comparison to NIST-tested algorithms.²⁹¹ Without access to proprietary source codes, there is little objective support that the underlying methodology is valid absent testing. Thus, untested FRT algorithms, like DataWorks Plus, used by the Detroit Police Department and the NYPD, do not meet the testability factor because there is no objective way to verify validity through reliable and consistent results.²⁹² A court would rely on a subjective assertion that DataWorks Plus’ methodology is valid, leaving inadequate support for effective cross-examination.

Alternatively, the NYPD and Detroit Police Department could argue that DataWorks Plus undergoes testing when the software is trained during development. However, in the Ninth Circuit reevaluation of *Daubert*, the court reasoned that scientific evaluations developed for

287. NISTIR, *supra* note 102, at 2.

288. *United States v. Gissantaner*, 990 F.3d 457, 464 (6th Cir. 2021); e.g., *United States v. Morgan*, 53 F. Supp. 3d 732, 741 (S.D.N.Y. 2014), *aff’d*, 675 F. App’x 53 (2d Cir. 2017) (highlighting no requirement to conduct the studies in all possible scenarios, just enough that a sound conclusion can be made).

289. See NISTIR 8271, *supra* note 102, at 44 (stating, “[d]espite migration to CNN-based technologies there is no consensus . . . This diversity of approaches[] suggests there is no prospect of a standard template[,] something that would require a common feature set to be extracted from faces.”).

290. CARL G. HEMPEL, *PHILOSOPHY OF NATURAL SCIENCE* 16 (1966); see *United States v. Shipp*, 422 F. Supp. 3d 762, 776 (E.D.N.Y. 2019); *United States v. Gissantaner*, 990 F.3d 457, 464 (6th Cir. 2021).

291. See, e.g., NISTIR 8280., *supra* note 30, at 71 (listing the patent number of an algorithm); Facial Recognition, U.S. Patent No. 8,798,336 (filed Sept. 23, 2013); *Clearview AI Principles*, CLEARVIEW AI, <https://www.clearview.ai/principles> [<https://perma.cc/PN65-CLEE>] (“Clearview AI’s proprietary technology processes the image and returns links to publicly available images that contain faces similar to the person pictured in the uploaded image.”).

292. NISTIR 8271, *supra* note 102, at 14.

litigation should tend toward inadmissibility because of concerns regarding impartiality and reliability.²⁹³ The court continued, “[i]f the proffered expert testimony is not based on independent research, the party proffering it must come forward with other objective, verifiable evidence that the testimony is based on ‘scientifically valid principles.’”²⁹⁴ Developer “testing” is not disinterested or impartial because it raises concerns surrounding a company’s pecuniary interest and testing of its product in an unregulated market. If an expert conducting research after being hired as a witness lacks objectivity, a developer who tests a product to sell it should as well.²⁹⁵ Additionally, the lack of regulation surrounding FRT development leaves no industry guardrails to support the impartiality of developer testing.²⁹⁶ Ultimately, developer training should not suffice because impartiality and reliability are in serious doubt rendering assertions of validity subjective without independent testing or source code. Further, developer training would be hard to challenge absent full disclosure of training methods or the source code, making effective cross-examination difficult. The courts have been clear that there is no way to show if a technology works without testing, so it should not be afforded “scientific status.”²⁹⁷

B. *Peer Review and Publication*

The cornerstone of peer review is the “relevant scientific community” weeding out junk science by carefully scrutinizing a theory.²⁹⁸ The relevant community for FRT is biometric and forensic scientists, similar to those assembled by the NIST, which provides a public forum for collaboration between scientists, industry, and

293. *Daubert v. Merrell Dow Pharms, Inc.*, 43 F.3d 1311, 1317 (9th Cir. 1995).

294. *Id.* at 1317–18.

295. *Id.*

296. *See supra* notes 67, 82–91 and accompanying text.

297. *Id.* at 463–64; *see supra* notes 194–196 and accompanying text (discussing how incomplete fingerprinting analysis undermines the reliability of the data); *see also* *United States v. Kime*, 99 F.3d 870, 883–84 (8th Cir. 1996) (rejecting eyewitness testimony because the current research is incomplete and insufficient to determine validity).

298. *Gissantaner*, 990 F.3d at 464; *United States v. Bonds*, 12 F.3d 540, 559 (6th Cir. 1993) 559.

academia.²⁹⁹ While some courts have cautioned skepticism of trade organizations, this concern does not apply to the NIST, which provides standards for several scientific fields, not just FRT.³⁰⁰ Additionally, the NIST conducts collaborative reviews using industry standards more closely resembling a “bona fide process” than a rubber stamp.³⁰¹ Critically, the NIST studies point out shortcomings and limitations in their methods and results, which further signifies meaningful review by a knowledgeable community.³⁰² Finally, several independent evaluations of facial analysis techniques and FRT programs supplement the work done by the NIST.³⁰³ There is little doubt this is the type of well-respected review of methodology that *Daubert* envisioned.

Peer reviews also help a court determine the degree of skepticism used in evaluating FRT, which should inform the assessment regarding BLM protestors.³⁰⁴ Specifically, some publications argue FRT is flawed when used on Black and Brown individuals, like BLM protestors.³⁰⁵ However, experts can disagree about test methodology and conclusions without disrupting the factor’s sufficiency.³⁰⁶ This is likely true for the contention that the application of FRT to identify people of color is misplaced.³⁰⁷ Nonetheless, it supports a skeptical glaze for evaluating FRT applied to BLM writ large.

Lastly, the FBI, Detroit Police Department, and the NYPD use a form of peer review to verify results; however, this analysis belongs in the

299. *Work with NIST*, NAT’L INST. OF STANDARDS & TECH., <https://www.nist.gov/about-nist/work-nist> [https://perma.cc/9UYS-DL9M] (detailing the numerous partnerships between NIST, industry, academia, and other government agencies).

300. *Gissantaner*, 990 F.3d at 464; *About NIST*, *supra* note 92 (is a world-renowned agency that has existed for over 100 years).

301. *Gissantaner*, 990 F.3d at 464.

302. *See supra* notes 198–200 and accompanying text.

303. *Bacci*, *supra* 36; *Nazah* *supra* 37; *Belhumer* *supra* 38.

304. *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579, 593 (1993) (reasoning peer review is important because it increases the likelihood of discovery of substantial flaws in methodology); *see also* Francisco J. Ayala & Bert Black, *Science and the Courts*, 81 AM. SCIENTIST 230, 238 (1993) (discussing the importance of peer review).

305. *Lohr supra* 137; *Crumpler supra* 137; John J. Howard, Yevgeniy B. Sirotnin, Jerry L. Tipton, *Quantifying the Extent to Which Race and Gender Features Determine Identity In Commercial Face Recognition Algorithms*, U.S. DEP’T. OF HOMELAND SEC. SCI. AND TECH. DIRECTORATE BIOMETRIC IDENTITY TECH. CTR. (May 2021).

306. *United States v. Hunt*, 464 F. Supp. 3d. 1252, 1257 (W.D. Okla. 2020).

307. *United States v. Shipp*, 422 F. Supp. 3d. 762, 777 (E.D.N.Y. 2019).

next *Daubert* factor discussed: controlling standards.³⁰⁸ Peer review of examiner results should be considered in the context of the system's accuracy because examination by a second examiner reduces subjectivity.³⁰⁹ Even so, some courts could see this type of peer review as further support for admissibility.³¹⁰

C. Rates of Error & Controlling Standards

1. Error rates

The FBI software and Clearview AI have a known error rate that is applicable in situations like January 6th. First, considering January 6th, the FBI asserts that the FRT algorithm it has used since 2019 is between 99.12% and 99.72% accurate.³¹¹ Notably, these error rates are achieved with high-quality photographs like those taken on the Capitol grounds from high-tech camera surveillance footage in well-lit conditions.³¹² Even if some of these images more closely resembled wild images with lesser quality or were searched against wild images, like those from social media, the best false negative rate is still .0071%.³¹³ While false negative rates remain relatively low, the false positive rate carries more weight because it translates to the ability of FRT to identify individuals accurately.³¹⁴ Although the FBI refuses to directly report false positives because the NIST tests its FRT software, false positive rates are known.³¹⁵ Specifically, in high-quality January 6th images, the false positive rates can be as low as 0.2%, depending on the developer.³¹⁶

308. NYPD PATROL GUIDE, *supra* note 65; DETROIT POLICE DEP'T, *supra* note 65.

309. See *supra* notes 202–201 and accompanying text.

310. See *supra* notes 202–201 and accompanying text.

311. *Hearings*, *supra* note 50 (statement of Kimberly J. Del Greco, Deputy Assistant Director, FBI) (noting this accuracy rate applies to the FRT software the FBI has used since 2019).

312. NISTIR 8271, *supra* note 102, at 7.

313. *Id.* at 199 (calculating the error rate when high-quality images are searched against wild photos).

314. *United States v. Mitchell*, 365 F.3d 215, 239 (3d Cir. 2004) (reasoning a high false negative rate may be unsatisfactory for law enforcement which is not as important as the potential for false accusations); accord *United States v. Harris*, 502 F. Supp. 3d 28, 39 (D.C. Cir. 2020).

315. See *supra* notes 119–122 and accompanying text.

316. See NISTIR 8271, *supra* note 102, at 8–9 (concluding aging is a significant source of error across many systems; while all faces age, drug use and age delaying or face altering surgery have an unknown impact that is still being quantified); *Mugshot*

Similarly, Clearview AI's accuracy rate for high-quality images with wild images is between .025% and .974%.³¹⁷

Notably, the differences between the quality of the probe photographs and the demographic compositions of the subjects from January 6th and BLM result in strikingly different FRT error rates.³¹⁸ First, degraded images like those common to BLM protests from cell phones and CCTV cameras decrease these accuracy rates. Specifically, when wild images are combined with the degraded image quality, the error rate increases two to three times higher than with high-quality images.³¹⁹ Next, considering the demographic compositions of the subjects from BLM—the error rates increase for people of color.³²⁰ When demographics shift from white males to Black and Brown individuals, false-positive rates increase by a factor of 10 to 100, resulting in false-positive rates from 0.5% to 34%.³²¹ While the *Shipp* court determined an error rate of approximately 2.2% was unacceptable, here, FRT can misidentify people of color in almost one in every three matches.³²² Notably, Clearview AI and DataWorks Plus have not participated in the NIST demographic tests, so given the range of error rates, they could be accurate or wholly unreliable.³²³

Implementation Guide: Photographic Considerations related to Facial Recognition Software and Booking Station Mugshots, FED. BUREAU INVESTIGATION (Apr. 25, 2013) (indicating image quality is affected by resolution, lighting, pose, angle, facial orientation, and obstructions such as hats, masks, and hair); see also NISTIR 8280, *supra* note 30, at 7–8, 63 (explaining false negatives occur because of low image quality or aging, and false positives occur because the two facial images are similar because the subjects are as twins, siblings, closely related, or doppelgangers).

317. NISTIR 8271, *supra* note 102, at 55.

318. See *supra* notes 108–114 and accompanying text.

319. NISTIR 8271, *supra* note 102, at 42.

320. For example, Polish men aged thirty-five to fifty produced false positives in 1 of 25,000 images, but Nigerian women over age sixty had false positives in 1 out of 35. PATRICK GROTH, NAT'L INST. OF STANDARDS & TECH., NISTIR 8429 (DRAFT), FACE RECOGNITION VENDOR TEST (FRVT) PART 8: SUMMARIZING DEMOGRAPHIC DIFFERENTIALS 4 (2022).

321. See *supra* notes 129–132 and accompanying text. See also *Gender Shades*, *supra* note 125 (confirming that false positives can be 34% higher in women of color than in lighter-skinned males). *E.g.*, NISTIR 8280, *supra* note 30, at 8; FISWG IDENTITY GROUND TRUTH, *supra* note 135, at 16.

322. *United States v. Shipp*, 422 F. Supp. 3d. 762, 778 (E.D.N.Y. 2019) (reasoning it was not a reasonable error rate because other forensic sciences have rates as low as 1 in 10 billion)

323. See NISTIR 8429, *supra* note 320 (searching for both produces no results).

Finally, in both January 6th and BLM applications, the error rates increase because testing data sets are significantly smaller than the operational databases.³²⁴ For example, the NIST identification test database size grew from 640,000 to 12 million high-quality photos, which caused the error rate to grow from 0.27% to 0.45%.³²⁵ This shows that the error rate is still low even with a larger database size, which is likely to translate to the FBI operational database because it is roughly 12% of the size of the test database.³²⁶ Alternatively, the Clearview AI database is .012% the size of the NIST database.³²⁷ Given the significant gap, it is more challenging to translate the error rate from testing to practice. Probably even more troubling is that with a more extensive NIST database of 18.27 million images, still a paltry .018% of Clearview AI, the error rate was from 0.5% to above 10% when demographics were considered.³²⁸ This dramatic jump in error rate amplifies the concerns with translating testing to the real world in the BLM scenario.

Ultimately, while Clearview AI has a known error rate, it was determined using images more common to January 6th, not BLM. This presents a problem because BLM protestors' demographics and database growth increase error rates calling into question the ability to quantify an error rate for Clearview AI. Ultimately, these factors render the known error rate untranslatable to BLM. Even where the error rates are known, they can be significantly higher than the 2.2% cited in *Shipp*.³²⁹ The gap between the known and unknown is too significant to be reliable for FRT systems used on BLM protestors. Alternatively, error rates are low and directly applicable to the January 6th images because of subject demographics and image quality. Even though

324. See Garvie et al., *supra* note 21, at 47, 50–51; NISTIR 8271, *supra* note 102, at 9–10.

325. See *supra* notes 129–132 and accompanying text.

326. See NISTIR 8271, *supra* note 102, at 12; U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-16-267, FACE RECOGNITION TECHNOLOGY: FBI SHOULD BETTER ENSURE PRIVACY AND ACCURACY 26, 29 (2016) (pointing out the FBI's test database contained 926,000 photos while the operational database contains thirty million photos).

327. Drew Hartwell, *Facial Recognition Firm Clearview AI Tells Investors it's Seeking Massive Expansion Beyond Law Enforcement*, WASH. POST, (Feb. 16, 2022, 12:47 PM), <https://www.washingtonpost.com/technology/2022/02/16/clearview-expansion-facial-recognition/>; see NISTIR 8271, *supra* note 102, at 20.

328. NISTIR 8271, *supra* note 102, at 1; see *supra* notes 129–132 and accompanying text.

329. *United States v. Shipp*, 422 F. Supp. 3d. 762, 778 (E.D.N.Y. 2019).

there are concerns with database growth for the FBI, there is a basis for cross-examination through existing testing and error rates which supports admissibility.

2. *Controlling standards*

There are significant differences in procedural standards, specialized training, and practical experience of the FBI, NYPD, and Detroit Police Department, which impact admissibility. First, standards and controls support admissibility by reducing subjectivity and ensuring error rates apply consistently across examiners and algorithms.³³⁰ The FBI requires its employees to adhere to a policy implementation guide that details restrictions and requirements for using FRT systems and best practices established by the NIST and FISWG.³³¹ These objective standards for verifying an identification ensure the accuracy of the FRT system is not undercut by the human examiner. Conversely, the NYPD and Detroit Police Department's FRT policies and procedures do little more than parrot the standard steps for performing a match.³³² Their policies have no references to NIST or FISWG documents that detail requirements, standards, or best practices for operating FRT systems and conducting facial comparisons.³³³ Here there are no objective standards, and the absence of meaningful controls points to subjectivity, which courts have cited with disapproval.³³⁴

In all circumstances, because of the lack of regulation of FRT standards and procedures are voluntary and self-imposed, which can

330. *Daubert*, 509 U.S. at 593; *see* *United States v. Williams*, 583 F.2d 1194, 1198 (2d Cir. 1978); *Pettus v. United States*, 37 A.3d 213, 224–45 (D.C. 2012) (relying on the fact that FBI document examiners are trained and employ national standards by ASTM International).

331. *Garvie et al.*, *supra* note 21, at 4, 50, 58, 66; R46586, *supra* note 19, at 7–8; *see supra* notes 92–96 and associated text.

332. *See* NYPD PATROL GUIDE, *supra* note 65; DETROIT POLICE DEP'T, *supra* note 65.

333. NYPD PATROL GUIDE, *supra* note 65; DETROIT POLICE DEP'T, *supra* note 65.

334. *Pettus v. United States*, 37 A.3d 213, 224–45 (D.C. 2012) (relying on the fact that FBI document examiners are trained and employ national standards by ASTM International); *see* *United States v. Shipp*, 422 F. Supp. 3d 762, 779–82 (E.D.N.Y. 2019) (determining the “near total subjectivity” of firearms-toolmark identification does not meet controlling standards and weighs against admissibility).

raise compliance questions.³³⁵ In response, the FBI self-audits FRT use by soliciting feedback and monitoring users.³³⁶ While not the form of independent auditing that accompanies accreditation or licensing, the FBI has been the subject of GAO audits, which provide some independent oversight of its internal processes.³³⁷ Similarly, New York recently passed a law requiring reporting and oversight of law enforcement FRT use.³³⁸ The New York statute is relatively new and might provide support in the future. But the fact remains the procedures themselves are lacking, and oversight of scant procedures does not cure subjectivity. Alternatively, Detroit does not have independent audits, but law enforcement self-reports FRT use statistics.³³⁹ Unlike the GAO audits, self-reported statistics do not directly substantiate that internal procedures are being followed; at best, it provides an indicator to warrant further investigation. Yet, evidence of false arrests shows the NYPD and Detroit used FRT as the sole method of identification directly against policy.³⁴⁰ This indicates that subjectivity sometimes rules the day. In the end, the deficient standards and lax controls of the NYPD and Detroit impermissibly skewed the reliable methods of facial analysis.³⁴¹

Additionally, standards and controls can mitigate cognitive bias, which creates reliability concerns for all suspects because of the nature

335. See *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579, 594–95 (1993); U.S. GOV'T ACCOUNTABILITY OFF., GAO-21-518, *FACIAL RECOGNITION TECHNOLOGY: FEDERAL LAW ENFORCEMENT AGENCIES SHOULD BETTER ASSESS PRIVACY AND OTHER RISKS* 20–21 (2021) (indicating there is no legal requirement for disclosure); see also *Facial Comparison Overview & Methodology Guidelines*, *supra* note 26, at 1, 4; R46586, *supra* note 19, at 15.

336. See *supra* notes 62–65 and accompanying text.

337. See *Facial Identification Subcommittee*, NAT'L INST. OF STANDARDS & TECH., <https://www.nist.gov/organization-scientific-area-committees-forensic-science/facial-identification-subcommittee> [<https://perma.cc/3PWL-R5XU>] (listing standards that have been completed and underdevelopment); *Certifications*, INT'L ASS'N FOR IDENTIFICATION, <https://www.theiai.org/certifications.php> [<https://perma.cc/JU9J-GEEM>] (listing certifications provided, which do not include facial comparisons or similar FRT techniques); see also *United States v. Beverly*, 369 F.3d. 516, 530 (6th Cir. 2004) (noting labs undertaking DNA forensic work are accredited by American Society of Crime Laboratory directors, an external agency).

338. Stop Hacks and Improve Electronic Data Security (SHIELD) Act, N.Y. GEN. BUS. LAW §§ 899-aa, 899-bb (McKinney 2020)

339. DETROIT POLICE DEP'T, WEEKLY REPORT ON FACIAL RECOGNITION (Dec. 7, 2020).

340. Fowler, *supra* note 5; See Allyn, *supra* note 66.

341. See *supra* notes 97–99.

of criminal investigations.³⁴² The FBI recently started to train individuals about the dangers of bias; however, it continues to ignore racial bias in FRT software.³⁴³ Ultimately, the FBI conducts double-blind evaluations and audits, providing review and oversight, which is likely adequate mitigation.³⁴⁴ Alternatively, it is unclear whether the NYPD or Detroit meaningfully addresses bias which is a substantial concern for Black and Brown individuals like BLM protestors. Specifically, increased surveillance and overrepresentation in search databases means law enforcement disproportionately applies FRT to Black and Brown populations.³⁴⁵ For example, the NYPD “gang affiliate” database is 99% Black and Latinx, and Detroit’s reports show 97% of its FRT searches were of Black and Brown individuals.³⁴⁶ These disparities demonstrate that a lack of internal processes allows systemic racial bias to continue, which can also provide a fertile environment for unreliable conclusions.³⁴⁷

Equally as important as standards and procedures for FRT use are the training and expertise of the examiners because FRT ultimately relies on human adjudication of recommended matches.³⁴⁸ The FBI requires all users to complete training consistent with the FISWG

342. See COMM. ON IDENTIFYING THE NEEDS OF THE FORENSIC SCIS. CMTY., NAT’L RSCH. COUNCIL, CONT. NO. 2006-DN-BX-0001, STRENGTHENING FORENSIC SCIENCE IN THE UNITED STATES: A PATH FORWARD, 122–24 (2009) (giving examples of the impacts various biases have, such as “contextual bias[es],” “anchoring,” and seeing patterns where there are none); see *supra* notes 140–144 and accompanying text (explaining cognitive bias is largely tied to mental shortcuts based on previous knowledge).

343. James Freeman, Opinion, *FBI Bias Training*, WALL ST. J. (June 19, 2018, 6:18 PM), https://www.wsj.com/articles/fbi-bias-training-1529445819?mod=nwsrl_cross_country&cx_refModule=nwsrl [https://perma.cc/MT2T-E8AS]; see Press Release, DOJ, Department of Justice Announces New Department-Wide Implicit Bias Training for Personnel (June 27, 2016), <https://www.justice.gov/opa/pr/department-justice-announces-new-department-wide-implicit-bias-training-personnel> [https://perma.cc/E3DJ-UZZX]; see *Hearings*, *supra* note 50, at 45 (statement of Kimberly J. Del Greco, Deputy Assistant Director, FBI) (“the system doesn’t look at skin tone and features [because it’s] a mathematical computation”).

344. See *supra* notes 148–150 and accompanying text.

345. Najibi, *supra* note 20; see *supra* Section I.A.3 (discussing the lack of mandatory regulation and spotty compliance by FBI in FRT).

346. DETROIT POLICE DEP’T, WEEKLY REPORT ON FACIAL RECOGNITION (Dec. 7, 2020) (report from the start of 2020 through December 6th, 2020, of the 114 total probe photos used for searching, 111 were of Black individuals).

347. See *supra* notes 140–148 and accompanying text.

348. R46586, *supra* note 19, at 5, 9–10, 17.

Guidelines and Recommendations for Facial Comparison Training to Competency.³⁴⁹ Additionally, the FBI conducts double-blind peer reviews of FRT matches which helps to bolster the expert's conclusion.³⁵⁰ In some instances, the more an evaluation relies on technical skill, the more an expert's extensive experience, use of standard procedures, or secondary review are essential factors.³⁵¹ The FBI's FACE Services Unit houses experts with training and experience, providing enough expertise to support this inquiry. The NYPD and Detroit Police Department might limit FRT to certain individuals allowing expertise to develop. However, these individuals are not trained and do not have standard procedures to support that "expertise." Consequently, even though the NYPD and Detroit both conduct peer reviews of matches, they are more likely to confirm an unreliable match.³⁵² Peer review cannot remedy the fact that untrained experts are performing the match—particularly when the data shows untrained individuals are more likely to misidentify individuals.³⁵³ Particularly in light of the high error rates for BLM, the lack of training and standards does not support admissibility.

D. General Acceptance

The *Daubert* court was clear, "[a] known technique which has been able to attract only minimal support within the community may properly be viewed with skepticism."³⁵⁴ Even with the widespread acceptance of facial comparison, courts should consider the history of law enforcement explicitly, indicating FRT is not a means of positive

349. See *Hearings, supra* note 50, at 4–5, 16, 25, 32, 41, 45 (statement of Kimberly J. Del Greco, Deputy Assistant Director, FBI) (explaining one officer submits the photo and a second specially trained officer completes the matching returning the reviewed results to the original requestor).

350. See *Hearings, supra* note 50, at 4–5, 16, 25, 32, 41, 45 (statement of Kimberly J. Del Greco, Deputy Assistant Director, FBI).

351. See *supra* notes 214–223 and accompanying text.

352. When examining the policies for both Police Departments, there is no mention of training. Conversely, the FBI is clear training is required, and only qualified individuals may use the FRT systems. NYPD PATROL GUIDE, *supra* note 65; DETROIT POLICE DEP'T, *supra* note 65; see *supra* note 96 and accompanying text.

353. Phillips *supra* 95.

354. *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579, 594 (1993).

identification.³⁵⁵ While FRT arises out of well-supported biometric principles of facial analysis, in novel applications like identification in a criminal trial, a court should be wary and apply additional scrutiny. Similarly, FRT applied to people of color does not enjoy widespread acceptance, which, combined with its novelty, warrants a heightened level of skepticism.³⁵⁶ In *United States v. Williams*³⁵⁷ and *Morgan*, the courts both applied increased scrutiny to novel iterations of DNA methodologies despite the widespread acceptance of traditional DNA evidence.³⁵⁸ In both cases, the court proceeded with caution, looking to the other factors for support which should also be the case for FRT.³⁵⁹ In sum, FRT is at best in the “twilight zone” due to a lack of consensus; thus, much like polygraph evidence, courts should be wary of “opening a legal Pandora’s box.”³⁶⁰

E. Rule 702: Relevancy & Fit

FRT, like other forensic evidence, would be used to place a suspect at a crime scene, which is resoundingly relevant for BLM and January 6th.³⁶¹ In the case of BLM—because the images come from CCTV and cellphones, it is more likely a jury would be unable to identify a suspect by simply viewing the images.³⁶² This further supports the relevancy of FRT because the identification assists the jury in determining a fact at

355. See *Hearings*, *supra* note 50, at 3–4 (statement of Kimberly J. Del Greco, Deputy Assistant Director, FBI) (declaring FRT cannot be used alone for positive identification and must be corroborated with additional evidence or witness identification); NYPD PATROL GUIDE: FACIAL RECOGNITION TECHNOLOGY (Mar. 12, 2020); DETROIT POLICE DEP’T, MANUAL DIRECTIVE 307.5 FACIAL RECOGNITION (Sept. 12, 2019).

356. Lohr *supra* 137; Crumpler *supra* 137; Howard *supra* 305.

357. 382 F. Supp. 3d 928 (N.D. Cal. 2019).

358. *Williams*, 382 F. Supp. 3d at 936-38 (examining a DNA mixture of five individuals); *United States v. Morgan*, 53 F. Supp. 3d 732, 735, 741 (S.D.N.Y. 2014), *aff’d*, 675 F. App’x 53 (2d Cir. 2017) (applying novel DNA modeling and testing methodology).

359. *Williams*, 382 F. Supp. 3d at 936-38; *Morgan*, 53 F. Supp. 3d at 735, 741.

360. *Frye v. United States*, 293 F. 1013, 1014 (D.C. Cir. 1923), *superseded by rule*, *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579 (1993); *see also* *Bonds*, 12 F.3d at 562 (reasoning “neither consensus nor certainty is needed, an absence of consensus is not immaterial”).

361. *United States v. Beasley*, 102 F.3d 1440, 1447 (8th Cir. 1996).

362. *See, e.g., United States v. Alexander*, 816 F.2d 164, 167–68 (5th Cir. 1987) (allowing expert testimony on identification from photograph to surveillance footage based on nose and mouth area, chin line, ear contours, etc. was proper).

issue not readily deciphered with their own knowledge.³⁶³ Alternatively, it is hard to see how FRT would benefit a jury when high-quality images or other identification evidence is available.³⁶⁴ In many cases, the images produced on the Capitol grounds and additional evidence on social media websites would allow the jury to identify the January 6th participants on their own.³⁶⁵ If a juror can make a common-sense determination without technical aid, FRT testimony should be “superfluous” and inadmissible.³⁶⁶

The second consideration for fit is whether the conclusion (an accurate identification by FRT) appropriately applies to the facts—in the case of January 6th, it does. First, from a FRT software perspective, the FBI system is validated by testing and has a known and reliable error rate for the image quality and demographics common to January 6th.³⁶⁷ Second, from a user perspective, the FBI has objective procedures for image evaluation and officer training that meets generally accepted industry standards.³⁶⁸ Cumulatively this results in a trained and experienced FBI expert using a tested system with a low error rate for the specific application. Further, the match is confirmed using a standard process and verified by a peer review. Based on these facts, conclusions drawn by the FBI would satisfy the fit factor because it is reasonably based on reliable principles and theories.³⁶⁹

However, applying FRT to the BLM protestors does not support a reliable conclusion because of differences in the system validity and accuracy and the expert’s training and procedures. First, from a FRT

363. *United States v. Sellers*, 566 F.2d 884, 886 (4th Cir. 1977) (allowing expert testimony on evaluation of photographs including effects of light, shadow, reflections, and distortions where the sole issue in case was the identity of a sole bank robber wearing a disguise).

364. See *Minor v. United States*, 57 A.3d 406, 423 (D.C. 2012) (discussing the numerous biases in the witness’s identification of the subject); *United States v. Smithers*, 212 F.3d 306, 317 (6th Cir. 2000) (conceding the court is more likely to admit expert testimony when the “eyewitness testimony was the crucial, if not the sole basis for Smithers’s conviction”).

365. *Day of Rage*, *supra* note 274.

366. *United States v. Kime*, 99 F.3d 870, 884 (8th Cir. 1996).

367. See *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579, 590–91 n.9 (1993) (emphasizing the difference between validity and reliability); NISTIR 8238, *supra* note 44, at 7, 9 (varying from “a few tenths of one percent up to beyond fifty percent,” changes in appearance from aging cause similarity scores to decline and accuracy progressively decreases).

368. See *supra* notes 92–96 and accompanying text.

369. *Gen. Elec. Co. v. Joiner*, 522 U.S. 136, 147 (1997).

software perspective, Detroit's DataWorks Plus software appears completely unvalidated through testing or source code evaluation, rendering the actual error rate unknown. Similarly, while the NYPD's other system Clearview AI has been tested and has a known error rate, it generally applies to high-quality images of less diverse subjects. Yet, testing shows that error rates increase for image quality and demographics common to BLM protestors.³⁷⁰ Second, from a user perspective, the NYPD and Detroit Police Department do not have objective procedures or controls that minimize human and system-level inaccuracies.³⁷¹ This all means an untrained officer with a high or unknown error rate using a subjective process verified results from an untested system. Such a conclusion is undoubtedly unreliable and not reasonable, especially in a criminal trial. Thus, for a BLM protestor, fit does not support admissibility.

F. The Admissibility of FRT for January 6th and BLM

Considering all factors, FRT is admissible for January 6th for the following reasons. First, general acceptance in the relevant community acknowledges that facial comparison techniques developed out of well-supported biometric methodologies.³⁷² Although the relevant community agrees that FRT results are not sufficient proof of identity, general acceptance only requires a "substantial portion of the pertinent scientific community," which could be achieved given the support for the underlying science of facial comparison.³⁷³ Second, numerous peer-reviewed articles and independent evaluations on facial comparison and analysis methods favor admissibility.³⁷⁴ While there is disagreement on the soundness of FRT software in certain applications, it does not apply to the image quality and demographics of the January 6th probe photos. Ultimately, peer review and the general acceptance of the methodology of facial comparison support admissibility. This underlying support should create a presumption of reliability that will color the evaluation of the remaining factors.

370. See *supra* notes 108–114 and accompanying text.

371. NISTIR 8280, *supra* note 30, at 7–8; see GAO-19-579T, *supra* note 53, at 19, 24–25, 42 (indicating GAO and the FBI disagreed about the 85% accuracy rate absent a false positive rate).

372. See *supra* Section II.D and Section I.A.1.

373. *United States v. Bonds*, 12 F.3d 540, 561 (1993); see *supra* notes 24–25, 355 and accompanying text. See *supra* Section II.D.

374. See *supra* Section II.B.

Next, considering the testability factor—the NIST tested the FBI FRT software in a black box study, which is widely accepted as the industry standard.³⁷⁵ While the NIST did acknowledge that discredited methods can be and probably are used given the wide range of accuracy, because the FBI software has been tested and performed accurately, it is unlikely in this case. Ultimately, the NIST independent verification of the FBI’s software supports testability and provides objective fodder for effective cross-examination.

Additionally, the NIST testing displays that the error rates for January 6th are near the apex of reliability because the factors that increase error rates (image quality and demographics) are less applicable, if at all.³⁷⁶ Even if image quality and test database size are issues, most algorithms maintain a sub-1% false positive rate for white males.³⁷⁷ Additionally, the FBI mitigates errors using clear standards adopted by the relevant community, training programs, specialized officers, audits, and peer reviews.³⁷⁸ These standards and controls reduce subjectivity, rendering little reason to believe the underlying sound science is undermined. Even if there were concerns, enough has been done to allow for effective cross-examination, so this factor favors admissibility for January 6th.

Ultimately, the fit factor is also satisfied in the case of January 6th. First, the validity of the methodology and application is corroborated by testing and supported by the relevant community and peer review process. Additionally, sophisticated FBI experts with clear procedures to reduce bias and subjectivity mitigate already low error rates for applications like January 6th. In light of these facts, the expert’s conclusion the match was accurate is reasonable. Finally, a court evaluates FRE Rule 403, balancing the highly persuasive nature of expert testimony, the relevancy to the jury, and the potential for undue prejudice.³⁷⁹ The relevancy requirement of FRE Rule 403 could render

375. *See supra* notes 285–288 and accompanying text.

376. *See supra* notes 108–114 and accompanying text (discussing the higher error rates for non-white, female, or older persons when compared to white, middle-aged men).

377. *See supra* notes 313–317 and accompanying text.

378. *See supra* Section II.C.2.

379. *United States v. Posado*, 57 F.3d 428, 435 (5th Cir. 1995) (reasoning FRE Rule 403 could play an “enhanced role for Rule 403 . . . particularly when the scientific or technical knowledge proffered is novel or controversial.”)

January 6th FRT evidence inadmissible.³⁸⁰ In many cases, the January 6th participants provided additional evidence on personal social media websites making FRT testimony “superfluous” and needlessly cumulative.³⁸¹ Next, it is hard to see how undue prejudice occurs in light of the *Daubert* evaluation because FRT is accurate and reliable for the January 6th scenario.

Alternatively, these same factors weigh against the admissibility of FRT for BLM. First, the peer review and general acceptance factors point to a skeptical glaze that should be applied to the BLM analysis. Importantly the relevant community agrees that some methods are not sound for forensic applications of facial comparison, and a wide variation in methods used is acknowledged.³⁸² Furthermore, peer-reviewed articles point to concerns with FRT algorithms’ ability to accurately identify people of color, like those participating in BLM protests.³⁸³ While both of these factors are generally met considering FRT overall, a court should not overlook the serious concerns with the validity of some algorithms and proceed with a skeptical eye because the concerns directly apply to BLM protestors, as detailed below.

First, DataWorks Plus has not been tested, which is crucial because it is one of the only ways to evaluate if FRT systems contain sound science since an algorithm’s source code is often proprietary.³⁸⁴ Suppose a court allowed the NIST testing to translate to the untested DataWorks Plus. In that case, it could masquerade as valid without being so because the absence of standard facial analysis methods means FRT algorithms could use discredited methods or theories. Again, the question is not in the method of the test but in the method tested, which is directly relevant to validity because only tested methodologies should be afforded scientific status. Without testing or access to the source code, there is no objective way to challenge DataWorks Plus through cross-examination.

380. See *Minor v. United States*, 57 A.3d 406, 423 (D.C. 2012) (commenting on the numerous biases in the witness’s identification of the subject); *United States v. Smithers*, 212 F.3d 306, 317 (6th Cir. 2000) (conceding the court is more likely to admit expert testimony when the “eyewitness testimony was the crucial, if not the sole basis for Smithers’s conviction”).

381. See *supra* note 364 and accompanying text.

382. See *supra* Section II.D.; *supra* notes 289–290 and accompanying text.

383. See *supra* notes 304–307 and accompanying text.

384. See *supra* notes 40–44 and accompanying text.

Second, the relatively high or unknown error rates and lack of controls and standards do not support admissibility. Specifically, the error rates are higher for BLM protestor demographics and low-quality images like cell phone video or CCTV.³⁸⁵ The known false positive rate for Black and Brown people can be up to 10%, but for wild images of Black and Brown people, it can increase up to three times.³⁸⁶ Given Clearview AI's operational database size, extrapolating from high and uncertain error rates reduces reliability even more.³⁸⁷ Further, the lack of specialized training or transparent criteria for determining a match does not mitigate a high error rate and might exacerbate it.³⁸⁸ Even conceding experience, it is not clear there is training, and there are examples of non-compliance.³⁸⁹ As is the case here, voluntary standards and subjectivity weigh against admissibility when the error rate is high, standards are absent, or non-compliance occurs.³⁹⁰ Consequently, this factor does not support admissibility.

Third, the conclusion that FRT accurately matched a BLM protestor does not fit the facts previously laid out. Even conceding the science and methodology of FRT are sound, there is too great a gap to apply that to Black and Brown individuals, particularly in low-quality images using software without testing. The lack of training and standards by the NYPD and the Detroit Police Department does little to mitigate high error rates. Further, a lack of testing and racial bias concerns create suspicion of the reliability of an expert's conclusion that the match was accurate. Ultimately, for BLM, the high error rate, lack of standards and testing, and tangible concerns surrounding racial bias cast doubt on the reliability of an expert's conclusion fitting the facts of the case.

Finally, for BLM FRE Rule 403, undue prejudice is significant because of the data supporting the disparate application of FRT on Black and Brown populations by the NYPD and Detroit Police Department.³⁹¹ This directly factors into the undue prejudice evaluation because FRT would act as a forensic eyewitness trusted by juries more than it should be while concealing concerns about

385. See *supra* notes 317–321 and accompanying text.

386. NISTIR 8271, *supra* note 102, at 42.

387. See *supra* notes 128–134 and accompanying text.

388. See *supra* notes 203–205 and accompanying text.

389. See *supra* notes 339–340 and accompanying text.

390. See *supra* notes 203–206 and accompanying text.

391. See *supra* Section II.C.1.

reliability, accuracy, and bias.³⁹² Importantly, juries believe forensic evidence is objective and impartial; thus, it is assumed to be reliable and validated.³⁹³ Consequently, FRE Rule 403 should operate as a last resort, rendering FRT inadmissible for BLM protestors.³⁹⁴

CONCLUSION

BLM and January 6th show the thin line between admissibility and exclusion for FRT evidence. For January 6th, the FBI uses independently tested FRT software, engendering confidence in the reported error rate and verifying the validity of the algorithm's methodology. Additionally, when the FBI applies that same FRT software to high-quality images of white males, the algorithms produce the most accurate results. Further, the training and experience of FBI experts, coupled with articulable standards for operating FRT and verifying matches, reduce subjectivity and bolster accuracy. Ultimately, it is a reasonable conclusion that FRT produced a reliable match for January 6th. Disagreements among experts do not defeat admissibility because the aforementioned factors provide facts to parse in cross-examination; thus, the adversarial process can identify shortcomings. Consequently, FRT is admissible because it is the jury's province to determine the weight of such evidence as *Daubert* mandates.³⁹⁵

Alternatively, the NYPD and Detroit Police Department use FRT software that has not been tested or tested in a different scenario from BLM. Untested algorithms can use discredited methods because of the lack of regulations. The wide range of accuracy across the industry directly calls into question untested algorithms basis on sound science. The untested DataWorks Plus should not be admissible on these facts alone because, without testing, only a subjective assertion remains. Clearview AI has been tested, but the testing did not use the racial composition and image quality common to BLM. The resulting error rates increase dramatically with degraded images of people of color. The NYPD and Detroit Police Department's lack of training and criteria for operating FRT exacerbate these error rates and undercut

392. See *supra* Section I.A.4 (discussing the numerous factors that cause errors in FRT).

393. Nakhaeizadeh et al., *supra* note 140, at 529.

394. See FED. R. EVID. 403 (stating evidence may be excluded when deemed dangerous due to "unfair prejudice" or risks "misleading the jury").

395. See *supra* note 248 and accompanying text.

sound scientific principles by injecting subjectivity. Together these factors mean it is unreasonable for an expert to conclude FRT software produced a reliable match in the case of BLM.

The balancing test of FRE Rule 403 also tips toward inadmissibility because there is a strong potential for undue prejudice. FRT expert testimony would be highly persuasive to the jury. At the same time, it is more likely to misidentify people of color, like those in BLM. Further, there is little basis for meaningful cross-examination due to the gap between sound science and subjectivity created by a lack of testing, error rates, and procedures. The subjectivity and unreliability of FRT coupled with FRE Rule 403 confirm the exclusion of FRT evidence for BLM protestors.

This outcome begs the question—should law enforcement use FRT on anyone if the industry’s lack of regulation and uniformity creates inconsistent results? Fortunately, the *Daubert* standard and the Federal Rules of Evidence provide flexibility by allowing courts to consider FRT in the context of specific factual scenarios as the software and industry continue to develop. Unfortunately, the fact remains, in the absence of federal regulations, the FBI polices itself, uses photos of individuals without informed consent, and largely ignores issues of bias and racial inequity in FRT systems and their application. Similarly, in most states, the patchwork of laws and local ordinances generally regulate the use of data by private companies or mandate disclosure but do not address law enforcement use or the underlying technology. Individuals may think they are protected because their police department cannot use FRT. Still, their driver’s license photo may very well be in the hands of the FBI and searchable by any number of police departments.

This begs another question—why is there no federal legislation in the face of the known challenges FRT creates? The state and local laws are a start, but some are being rolled back, with elected officials citing the crime rates as justification.³⁹⁶ But does allowing law enforcement to use unregulated technology with little oversight make us safer? We would be wise not to let a perceived threat convince us we do not need to or should not legislate police use of FRT in the name of increased security. While the question of admissibility is one for the courts, the question of use is one of policy that we all need to consider before the decision is made for us.

396. Dave, *supra* note 88.