SoundThinking's Black-Box Gunshot Detection Method: Untested and Unvetted Tech Flourishes in the Criminal Justice System

Brendan Max*

26 STAN. TECH. L. REV. 193 (2023)

ABSTRACT

SoundThinking has successfully marketed their ShotSpotter forensic gunshot detection method to police departments and prosecutors as a reliable method for detecting and locating gunfire incidents in urban environments and generating admissible evidence for use in criminal prosecutions. The ShotSpotter method involves networks of microphones deployed in urban settings, which are tasked with detecting the impulsive sounds of gunfire and transmitting the sound recordings to SoundThinking's black-box algorithm and human examiners for forensic analysis. If the impulsive noises are suspected to have originated from gunfire, SoundThinking alerts local police departments so that officers can quickly respond and investigate. SoundThinking promotes ShotSpotter as a high-tech improvement on traditional 911 calls. The reliability of ShotSpotter is hindered by the technically-challenging environments where ShotSpotter systems are deployed (neighborhoods with dense buildings and other infrastructure) and the routine occurrence of impulsive noises (from vehicle traffic, construction equipment, and many other sources) that are known to trigger ShotSpotter false alerts. To assess the ability of methods like SoundThinking's gunshot detection method to reliably complete their forensic tasks and to quantify important rates of error, method developers like SoundThinking are supposed to engage in a multi-step development process involving validation testing, algorithm verification, and error rate analysis. Yet SoundThinking has largely ignored this development process, instead promoting accuracy and performance claims that have no legitimate scientific bases. And neither the scientific community nor the judicial system have engaged in the forms of oversight that should preclude the use of such untested forensic

* Chief, Forensic Science Division, Cook County Public Defender Office.

evidence in the criminal justice system. The result has been a proliferation of ShotSpotter systems in 150 U.S. cities and the use of ShotSpotter evidence in over 200 criminal trials without proof that the ShotSpotter method works and in the face of growing evidence that the method is plagued by a high incidence of false alerts. In light of the facts that ShotSpotter systems are deployed primarily in communities of color and the harms associated with ShotSpotter false alerts are borne primarily by people of color, both the scientific and legal communities need to engage in rigorous oversight of SoundThinking's forensic method. Until such oversight establishes the reliability of ShotSpotter evidence and its true rates of error, ShotSpotter evidence should play no role in the criminal justice system.

TABLE OF CONTENTS

INTRODUCTION

Michael Williams spent a year in jail—falsely accused of murder—because of a secret black-box[1] forensic method developed by SoundThinking.[2] In Mr. Williams' case, police and prosecutors relied on claims by SoundThinking that their ShotSpotter technology can reliably detect the sound of gunfire in complex urban environments, distinguish the sound of gunfire from all other manner of impulsive noises, and quickly and accurately lead police to the precise location of the suspected gunfire.[3] Based on these claims, police and prosecutors pursued flawed murder charges against Mr. Williams even though no witnesses implicated him, no physical evidence connected him to the murder, and no motive existed for Mr. Williams to have committed the murder.[4] Similar criminal prosecutions occur across the United States in more than 150 cities,[5] initiated by real-time alerts from SoundThinking to police departments that purport to conclusively report the occurrence and location of gunfire.

The rapid adoption of ShotSpotter evidence by police and prosecutors across the United States has occurred despite clear signs that SoundThinking's gunshot detection technology is flawed and routinely alerts police and prosecutors to locations where ShotSpotter has mistaken innocent environmental noises for gunfire. A growing body of data shows that ShotSpotter is routinely fooled into issuing false alerts to police agencies by noises that originate from a plethora of innocent environmental sources such

---

[1] Neil Savage, *Breaking into the Black Box of Artificial Intelligence*, NATURE (Mar. 29, 2022), https://perma.cc/VDZ5-RTPT (defining a black-box algorithm as a computer-based decision model where "researchers and users typically know the inputs and outputs, but it is hard to see what's going on inside" as the computer assesses inputs and reaches decisions).

[2] As a publicly traded for-profit company, SoundThinking is an outlier in the forensic science community. The large majority (199 of 212) of accredited forensic laboratories that provide evidence for use in the criminal justice system are government operated rather than private for-profit entities. None are publicly traded companies. For further information about accredited crime laboratories in the United States, visit https://search.anab.org/ (within "Scope of Accreditation" section, under the "Forensic Testing" tab, select "Any"; within "Programs/Schemes" section, within the "Forensic Testing and Inspection Programs" category, select "FBI QAS – Testing"; run search) (results as of May 24, 2023).

[3] *ShotSpotter Frequently Asked Questions*, SOUNDTHINKING, https://perma.cc/DU6Y-Z5Q2.

[4] Garance Burke et al., *How AI-Powered Tech Landed Man in Jail with Scant Evidence*, U.S. NEWS & WORLD REP. (Mar. 5, 2022), https://perma.cc/SVW4-HQW4 (reporting that ShotSpotter evidence "anchored the prosecutor's theory that Williams shot Herring inside his car, even though the case supplementary report from police did not cite a motive, nor did it mention any eyewitnesses. There was no gun found at the scene of the crime.").

[5] *ShotSpotter Frequently Asked Questions*, *supra* note 3.

as vehicle traffic, construction sounds, and other daily sources of impulsive noises.[6] While the high incidence of ShotSpotter false alerts has caused some police departments to discontinue the use of ShotSpotter, other police departments who continue to rely on ShotSpotter find that officers discover no evidence of real gunfire events during 89% of ShotSpotter-initiated investigations.[7] Despite these emerging signs of the unreliability of ShotSpotter technology, SoundThinking's reach into the criminal justice system continues to expand—prosecutors across the United States pursue criminal cases built around ShotSpotter evidence and criminal court judges routinely authorize the use of ShotSpotter evidence without the technology undergoing the type of scientific and legal vetting that theoretically should precede the admission of novel forensic evidence in criminal trials.

The quiet proliferation of unvetted ShotSpotter systems follows a clear trend involving the expanding reliance on black-box algorithms across many sectors of society. Generally defined as computer models that derive judgments from analyses of data in ways that make it hard or impossible for outside observers to reconstruct how the judgments were reached,[8] algorithms quietly work in the background of modern life for purposes as varied as influencing which online content Instagram users will be inundated with[9] to diagnosing cancer.[10] Likewise, demand for the use of computer algorithms at every step in the criminal justice system is growing at unprecedented levels.[11] Police now use

---

[6] *See infra* notes 55-63.

[7] *See infra* notes 172-77 and accompanying text.

[8] Jeremy Petch et al., *Opening the Black Box: The Promise and Limitations of Explainable Machine Learning in Cardiology*, 38 CANADIAN J. CARDIOLOGY 204, 204 (2022) (defining black-box algorithms as predictive models "that are sufficiently complex that they are not straightforwardly interpretable to humans").

[9] Adam Mosseri, *Shedding More Light on How Instagram Works*, INSTAGRAM (June 8, 2021), https://perma.cc/VNZ3-BTQD (explaining that Instagram uses multiple algorithms to assess user signals—browsing habits, interactions with other users, and the content of user posts—and then assesses thousands of such signals for each user and bases content recommendations on this signal analysis).

[10] Andre Esteva et al., *Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks*, 542 NATURE 115, 115 (2017) (reporting on an algorithm developed to detect skin cancer, concluding that the algorithm is "capable of classifying skin cancer with a level of competence comparable to dermatologists," and envisioning a time when such skin cancer diagnosis requires only a smartphone application).

[11] Michael Brenner et al., *Constitutional Dimensions of Predictive Algorithms in Criminal Justice*, 55 HARV. C.R.-C.L. L. REV. 267, 268 (2020) (noting that "[a]rtificial intelligence and algorithmic tools are rapidly becoming embedded in our criminal justice system"); DANIELLE KEHL ET AL., ALGORITHMS IN THE CRIMINAL JUSTICE SYSTEM: ASSESSING THE USE OF RISK ASSESSMENTS IN SENTENCING 3 (Berkman Klein Ctr. for Internet & Soc'y, Harv. L. Sch. 2017),

black-box algorithms to decide when and where to deploy police resources,[12] to predict who is likely to fall victim to gunshot crimes,[13] and to justify stopping, searching, and arresting people.[14] Crime lab personnel use computer algorithms for interpreting the meaning of DNA results[15] and searching fingerprint databases to try to identify the source of latent prints recovered from crime scenes.[16] Prosecutors use algorithms to determine the seriousness of charges to file against suspects[17] and to provide the justification for juries to vote in favor of conviction.[18] And judges use algorithms for critical

https://perma.cc/5HDA-5YA3 ("The rapid and unprecedented rise of predictive algorithms has been fueled by a number of factors, including the vast amounts of data generated by ubiquitous use of the internet and smart devices and a growing emphasis on data-driven decision-making in both our private lives and public policy. Unsurprisingly, this emphasis on the use of data in government has permeated many stages of the criminal justice system as well, from predictive policing to risk assessment in the corrections system.").

[12] Sarah Brayne & Angèle Christin, *Technologies of Crime Prediction: The Reception of Algorithms in Policing and Criminal Courts*, 68 SOC. PROBS. 608, 609 (2021) ("Over recent decades, the U.S. criminal justice system has witnessed a proliferation of algorithmic technologies. Police departments now increasingly rely on predictive software programs to target potential victims and offenders and predict when and where future crimes are likely to occur." (citation omitted)).

[13] Monica Davey, *Chicago Police Try to Predict Who May Shoot or Be Shot*, N.Y. TIMES (May 23, 2016), https://perma.cc/2TQZ-CC7Y**.**

[14] JOSEPH M. FERGUSON & DEBORAH WITZBURG, THE CHICAGO POLICE DEPARTMENT'S USE OF SHOTSPOTTER TECHNOLOGY 19 (Chi. Off. of Inspector Gen. 2021), https://perma.cc/VD5U-8JMM (reporting that in addition to thousands of police responses to individual ShotSpotter alerts, "[a]t least some officers, at least some of the time, are relying on ShotSpotter results in the aggregate to provide an additional rationale to initiate [a] stop or to conduct a pat down once a stop has been initiated") [hereinafter OIG REPORT]; Kashmir Hill, *Wrongfully Accused by an Algorithm*, N.Y. TIMES, https://perma.cc/474N-QWVK (Aug. 3, 2020) (reporting on the wrongful arrest of a man for theft based on flawed facial recognition evidence).

[15] Peter Gill et. al., *A Review of Probabilistic Genotyping Systems:* EuroForMix*,* DNAStatistX*, and* STRmix™, 12 GENES, no. 10, 2021, at 1 ("The use of software to evaluate DNA profile evidence is widespread in the forensic biology community.").

[16] Philip J. Kellman et al., *Forensic Comparison and Matching of Fingerprints: Using Quantitative Image Measures for Estimating Error Rates Through Understanding and Predicting Difficulty*, 9 PLOSONE, no. 5, at 2 (2014) ("It is common for a latent print to be submitted to an AFIS (automated fingerprint identification system) database, where automated routines return a number of most likely potential matches.").

[17] Jocelyn Gecker, *San Francisco Prosecutors Turn to AI to Reduce Racial Bias*, WASH. POST (June 12, 2019, 6:25 PM EDT), https://perma.cc/9X94-XKST (reporting that prosecutors in San Francisco worked with data scientists and engineers to implement an algorithm-based method for determining the appropriate level of criminal charges to pursue against defendants).

[18] *Survey Shows STRmix Has Been Used in 220,000 Cases Worldwide*, STRMIX (Nov. 19, 2020, 9:00 AM), https://perma.cc/7T88-XFH5 (reporting that DNA interpretation results from just one forensic DNA algorithm have "been used in at least 220,000 cases worldwide since its introduction in 2012").

determinations such as setting bond amounts[19] and selecting the length of incarceration at sentencing.[20]

But just as the rapid adoption of algorithm-based technology in broader society has sometimes resulted in unforeseen harm, so has the uncritical adoption of black-box technology in the criminal justice system. Not all black-box forensic methods are created equal. Rather, experienced coders regularly make basic mistakes when attempting to write reliable computer code.[21] In addition to coding errors, black-box algorithms can suffer from other developmental failures, including inadequate training data and the replication of human biases into algorithm operations and decision-making.[22] And as the use of algorithms has proliferated, so have the examples of algorithm failure, resulting in substantial societal harm from sources as varied as self-driving cars to computer-graded college exams.[23] For black-box algorithms employed in the

---

[19] Tom Simonite, *Algorithms Were Supposed to Fix the Bail System. They Haven't*, WIRED (Feb. 19, 2020, 8:00 AM), https://perma.cc/6JTC-MZVC (reporting on proposals to abandon the use of an algorithm-based method to determine bail amounts in New Jersey because algorithm bond determinations are "often built on data that reflects racial and ethnic disparities in policing, charging, and judicial decisions").

[20] State v. Loomis, 881 N.W.2d 749, 761 (Wis. 2016) (where a defendant objected to the use of an algorithm-based determination of his risk of reoffending during sentencing, arguing in part that his due process rights were violated when the judge based sentencing decisions on output from proprietary software).

[21] Christian F. Chessman, Note, *A 'Source' of Error: Computer Code, Criminal Defendants, and the Constitution*, 105 CALIF. L. REV. 179, 186 (2017) (discussing a study that found that 33% of highly experienced C++ coders failed to correctly use parentheses when coding basic equations, resulting in faulty source code).

[22] Thomas C. Redman, *If Your Data Is Bad, Your Machine Learning Tools Are Useless*, HARV. BUS. REV. (Apr. 2, 2018), https://perma.cc/C2L7-PUB4 ("Poor data quality is enemy number one to the widespread, profitable use of machine learning. . . . To properly train a predictive model, historical data must meet exceptionally broad and high quality standards. First, the data must be right: It must be correct, properly labeled, de-deduped, and so forth. But you must also have the *right* data — lots of unbiased data, over the entire range of inputs for which one aims to develop the predictive model."); Nicol Turner Lee, Paul Resnick & Genie Barton, *Algorithmic Bias Detection and Mitigation: Best Practices and Policies to Reduce Consumer Harms*, BROOKINGS INST. (May 22, 2019), https://perma.cc/R2AA-3LLW ("In the pre-algorithm world, humans and organizations made decisions in hiring, advertising, criminal sentencing, and lending. . . . Today, some of these decisions are entirely made or influenced by machines whose scale and statistical rigor promise unprecedented efficiencies. . . . However, because machines can treat similarly-situated people and objects differently, research is starting to reveal some troubling examples in which the reality of algorithmic decision-making falls short of our expectations. Given this, some algorithms run the risk of replicating and even amplifying human biases, particularly those affecting protected groups.").

[23] *See, e.g.*, Andrea Roth, *Machine Testimony*, 126 YALE L.J. 1972, 1994-95 (2017) ("Therac-25, a computer-controlled radiation therapy machine . . . 'massively overdosed' six people in the late 1980s based on a software design error."); Lauren Aratani, *Tesla Investigation*

criminal justice system, algorithmic error is likewise not rare—errors have occurred with DNA algorithms,[24] breathalyzer machines,[25] and risk-assessment algorithms used by judges when making sentencing decisions in criminal cases.[26] Even algorithms for relatively simple applications in the criminal justice

Deepens After More Than a Dozen U.S. 'Autopilot' Crashes, GUARDIAN (June 9, 2022, 2:53 PM EDT), https://perma.cc/6VYH-LUWD (reporting that the NHTSA has investigated numerous crashes of Tesla cars while operating during computer-assisted driving); Jesse Halfon, *Uber's Self-Driving Car Killed Someone. Why Isn't Uber Being Charged?*, SLATE (Oct. 20, 2020, 9:00 AM), https://perma.cc/WSA3-G4UR (reporting that a self-driving Uber car killed a pedestrian when the algorithm controlling the car operation mistook a woman for an inanimate object and overrode sensors that detected her presence six seconds before she was struck and killed); James Gleik, *Little Bug, Big Bang*, N.Y. TIMES MAG. (Dec. 1, 1996), https://perma.cc/MH82-3XA2 (discussing coding error that led to the crash of the $7 billion Ariane 5 rocket); Tom Simonite, *Meet the Secret Algorithm That's Keeping Students Out of College*, WIRED (July 10, 2020, 7:00 AM), https://perma.cc/2WLM-CAFZ (reporting that some students lost college scholarships when the International Baccalaureate program used a faulty computer algorithm to grade student performance); Eric Griffith, *10 Embarrassing Algorithm Fails*, PC MAG. (Sept. 23, 2017), https://perma.cc/97TX-8HWE (reporting that Tesla issued upgrades to their autonomous driving car computer code after a Tesla owner crashed into a tractor-trailer while in semi-autonomous mode); Julia Angwin et al., *Facebook Enabled Advertisers to Reach 'Jew Haters,'* PROPUBLICA (Sept. 14, 2017, 4:00 PM EDT), https://perma.cc/H9VC-SJN8 (reporting how a flaw in Facebook's algorithm allowed advertisers to locate and direct sales pitches through search terms such as "Jew hater" and "How to burn Jews"); Sophie Bushwick, *How NIST Tested Facial Recognition Algorithms for Racial Bias*, SCI. AM. (Dec. 27, 2019), https://perma.cc/V6BZ-WCLB (reporting that scientists at the National Institute of Standards and Technology tested 189 facial recognition algorithms and found that many of the algorithms were biased against people of color—producing significantly more accurate results when tested on Caucasian faces than on Black faces); Ramona Pringle, *When Algorithms Go Bad: Online Failures Show Humans Are Still Needed,* CBC NEWS, https://perma.cc/BN22-SYWY (Oct. 1, 2017) (reporting that when Amazon users innocently selected certain items for purchase, the Amazon algorithm unexpectedly offered users the chance to purchase the additional items that could be combined to make home-made explosive devices); Vincent Manancourt, *UK to End Controversial Visa Screening Algorithm*, POLITICO (Aug. 4, 2020, 12:57 PM CET), https://perma.cc/DPD3-FBZH (reporting that the United Kingdom discontinued the algorithm used to screen visa applications when it was discovered to discriminate based on nationality).

[24] David Murray, *Queensland Authorities Confirm 'Miscode' Affects DNA Evidence in Criminal Cases*, COURIER-MAIL (Mar. 20, 2015), https://perma.cc/3AXM-3KA8 (reporting that an error with a DNA interpretation algorithm resulted in incorrect interpretations in at least sixty cases); State v. Pickett, 246 A.3d 279, 296 (N.J. Super. Ct. App. Div. 2021) (documenting coding flaws in a different DNA interpretation algorithm which caused the system to "overestimate the likelihood of guilt").

[25] 246 A.3d at 296 (summarizing expert testimony regarding "thousands of faults discovered in the source code of breathalyzer systems").

[26] Julia Angwin et al., *Machine Bias*, PROPUBLICA (May 23, 2016), https://perma.cc/F5RT-9PGK (reporting that computer algorithms used by criminal judges to assess the risk of recidivism during sentencing procedures disproportionately classify Black people as higher risks for recidivism).

system, such as one that was supposed to accurately calculate release dates for people incarcerated in prison, are error prone.[27]

Due to the potential of coding and other problems,[28] which can be dormant and difficult to diagnose with black-box algorithms,[29] the rigorousness of the development process and the validity of resulting empirical performance data make all the difference. They provide the means for assessing scientific performance of forensic algorithms[30] and their appropriateness for use in criminal trials.[31] Unfortunately, the vetting of the development process for

---

[27] Dell Cameron, *Software 'Bug' Keeps Arizona Prisoners Behind Bars Past Release Dates*, GIZMODO, https://perma.cc/VK7S-H4FL (Feb. 22, 2021, 3:00 PM) (reporting that software used by the Arizona Department of Corrections was so riddled with bugs that it failed to properly calculate detainee release dates, resulting in hundreds of people being imprisoned past their eligible release dates).

[28] Lee, Resnick & Barton, *supra* note 22 (reporting that computer algorithms can fail for several reasons in addition to coding problems, including the introduction of historical human bias and incomplete or unrepresentative algorithm training data).

[29] INST. OF ELEC. & ELECS. ENG'RS, ETHICALLY ALIGNED DESIGN: A VISION FOR PRIORITIZING HUMAN WELL-BEING WITH AUTONOMOUS AND INTELLIGENT SYSTEMS 29 (2017), https://perma.cc/J5AJ-TK47 (reporting that "[algorithms] will be performing tasks that are far more complex and have more effect on our world than prior generations of technology. This reality will be particularly acute with systems that interact with the physical world, thus raising the potential level of harm that such a system could cause. . . . At the same time, the complexity of [algorithmic] technology will make it difficult for users of those systems to understand the capabilities and limitations of the [algorithms] that they use, or with which they interact.").

[30] Geoffrey Stewart Morrison & William C. Thompson, *Assessing the Admissibility of a New Generation of Forensic Voice Comparison Testimony*, 18 COLUM. SCI. & TECH. L. REV. 326, 363 (2017) (stating that "[e]mpirical testing of validity and reliability is the only way to demonstrate how well a forensic analysis system actually works"); Daniel C. Murrie et al., *Perceptions and Estimates of Error Rates in Forensic Science: A Survey of Forensic Analysts*, 302 FORENSIC SCI. INT'L, Sept. 2019, at 1 ("All scientific procedures inevitably involve some error, particularly when the procedures rely on subjective human judgment. Understanding the nature and extent of error is crucial to understanding the meaning of any particular piece of scientific evidence. Thus, most scientific disciplines work to document the reliability (i.e., consistency, reproducibility) and validity (i.e., accuracy) of their scientific procedures.").

[31] Murrie et al., *supra* note 30, at 1 ("For scientific evidence to be admitted in court, data regarding reliability and validity are crucial . . . A fact-finder's ability to properly interpret the conclusions of any scientific analysis depends upon the ability to know the chance that an error occurred."); Morrison & Thompson, *supra* note 30, at 379 ("In considering the admissibility of expert testimony, the *Daubert* ruling instructs the trial judge to consider 'whether the reasoning or methodology underlying the testimony is scientifically valid and . . . whether that reasoning or methodology properly can be applied to the facts in issue.' It goes on to state that 'a key question to be answered in determining whether a theory or technique is scientific knowledge that will assist the trier of fact will be whether it can be (and has been) tested . . . [T]he statements constituting a scientific explanation must be capable of empirical test.' Later it states that 'in the case of a particular scientific technique, the court ordinarily should consider the known or potential rate of error.' We interpret these statements as requiring the forensic scientist to empirically test the degree of validity and reliability of their system and provide the results of such tests to the judge so that the judge

algorithms like SoundThinking's poses challenges that the criminal justice system has not met. For decades, the criminal justice system has failed to provide meaningful vetting for a wide range of traditional (i.e., non-algorithm based) forensic methods,[32] leading to both the use of unreliable forensic evidence, such as bitemarks, hair comparison, bullet lead analysis, and others,[33] as well as the misuse of other types of forensic evidence.[34] The courts' task in

---

can take them into consideration when deciding on admissibility." (quoting Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579, 592-94 (1993))).

[32] *See, e.g.*, Erin Murphy, *Neuroscience and the Civil/Criminal* Daubert *Divide*, 85 FORDHAM L. REV. 619, 621, 624 (2016) (stating that "evidence proffered by plaintiffs in civil cases receives harsh scrutiny for reliability, whereas evidence proffered by prosecutors in criminal cases typically gets a free pass" and "nearly all of the common forensic techniques . . . routinely admitted by courts[] have been repeatedly denounced as lacking in any scientific basis"); Jane C. Moriarty, *Will History Be Servitude? The NAS Report on Forensic Science and the Role of the Judiciary*, 2010 UTAH L. REV. 299, 315 (2010) (reporting that "[i]n civil cases, courts seem quite up to the task of evaluating microbiology, teratology, and toxicology evidence," but "when it comes to evaluating the shortcomings of lip prints and handwriting, courts are unable to muster the most minimal grasp of why a standardless form of comparison might lack evidentiary reliability"); Aliza B. Kaplan & Janis C. Puracal, *It's Not a Match: Why the Law Can't Let Go of Junk Science*, 81 ALBANY L. REV. 895, 927 (2018) ("Courts continue to admit expert testimony based on bite mark analysis, firearms analysis, footwear analysis, fingerprint analysis, and toolmark analysis—each of which has been shown to lack scientific validity or be of only limited probative value."); Michael J. Saks & David L. Faigman, *Failed Forensics: How Forensic Science Lost Its Way and How It Might Yet Find It*, 4 ANN. REV. L. & SOC. SCI. 149, 161 (2008) ("The courts, and the legal profession more generally, have been remarkably ineffective in policing forensic science. With only a few notable exceptions, courts have not closely evaluated the scientific bases for forensic identification and have not been responsible for exposing the flaws in any of the forensic science fields that have, over time, been abandoned owing to their lack of validity. . . . [W]hen it comes to science, and particularly statistics, judges pause and sputter, wondering whether it is truly part of their responsibility to know the details of scientific methods."); D. Michael Risinger, *Navigating Expert Reliability: Are Criminal Standards of Certainty Being Left on the Dock?*, 64 ALBANY L. REV. 99, 104-08 (2006) (reviewing decades of published judicial opinions and finding that courts apply a different level of admissibility analysis in criminal cases than in civil cases, leading to criminal defendants losing more admissibility challenges when confronting prosecution scientific evidence and leading to scientific evidence proffered by criminal defendants being excluded more frequently).

[33] Connor Lynch, *The Problem with Forensic Sciences*, DISCOVER MAG. (Nov. 8, 2022, 2:00 PM), https://perma.cc/GR5S-CU3Z ("Forensic sciences, in general, have encountered serious problems with the basic assumptions that underlie individual techniques. . . . Forensics has more than its fair share of 'junk science,' which refers to faulty scientific information or research. Historically, the actual scientific basis for investigative techniques comes long after they've started to be used, if it comes at all.").

[34] Naomi Elster, *How Forensic DNA Evidence Can Lead to Wrongful Convictions*, JSTOR DAILY (Dec. 6, 2017), https://perma.cc/C9HB-35GU (describing problems with current forensic DNA analyses, including the sensitivity of testing methods, the susceptibility to contamination, the prevalence of false partial-profile matches in DNA databases, and confusion about the appropriate weigh to give interpretive opinions); PRESIDENT'S COUNCIL OF ADVISORS ON SCI. & TECH., FORENSIC SCIENCE IN CRIMINAL COURTS: ENSURING SCIENTIFIC VALIDITY OF

vetting black-box forensic evidence is even more challenging due to the fact that algorithmic interpretive decisions are obscured by complicated and proprietary computer code rather than offered only by human examiners who can be questioned under oath.[35] The courts' task is also critical because "defendants cannot see, understand, or challenge the findings."[36] Yet the criminal justice system has largely abdicated responsibility for such vetting,[37] paving the way for the uncritical use of ShotSpotter evidence in the criminal justice system.

The uncritical adoption of ShotSpotter gunshot detection technology is especially concerning due to the disparate impact that the technology has on communities of color. Across the United States, ShotSpotter systems are deployed primarily in communities of color.[38] When coupled with the routine

---

FEATURE-COMPARISON METHODS 87 (2016), https://perma.cc/NRB8-4MYQ (reporting that the forensic fingerprint comparison method "was long hailed as infallible, despite the lack of appropriate studies to assess its error rate"); ABS GROUP, ROOT AND CULTURAL CAUSE ANALYSIS OF REPORT AND TESTIMONY ERRORS BY FBI MHCA EXAMINERS 216 (2018), https://perma.cc/KZN3-KZXR (reporting a root cause analysis of the systematic misstatement of forensic hair comparison evidence by FBI examiners and documenting errors in "almost half" of expert reports reviewed and "[o]ver 90%" of expert testimonies reviewed).

[35] Yavar Bathaee, *The Artificial Intelligence Black Box and the Failure of Intent and Causation*, 31 HARV. J. L. & TECH. 889, 892-93 (2018) ("Humans can be interviewed or cross-examined; they leave behind trails of evidence such as e-mails, letters, and memos that help answer questions of intent and causation; and we can draw on heuristics to help understand and interpret their conclusions. If an AI program is a black box, it will make predictions and decisions as humans do, but without being able to communicate its reasons for doing so. . . . [L]ittle can be inferred about the intent or conduct of the humans that created or deployed the AI, since even they may not be able to foresee what solutions the AI will reach or what decisions it will make[.]").

[36] Brenner et al., *supra* note 11, at 278; *see also* KEHL ET AL., *supra* note 11, at 28 (stating that the nature of algorithms makes it "difficult for researchers and outside experts to evaluate and audit the algorithms in order to test for accuracy and bias. The lack of information about how inputs are weighted also makes it harder to bring legal challenges to the use of these tools[.]")

[37] See *infra* notes 204-16 for a discussion of the failure of trial court judges to rigorously apply legal admissibility standards to prosecution-proffered forensic evidence.

[38] Brief for Chicago Community-Based Organizations Brighton Park Neighborhood Council et al. as Amici Curiae Supporting Defendant at 14, State v. Williams, No. 20cr0899601 (Ill. Cir. Ct. Cook Cnty. May 2021), *available at* https://perma.cc/3WQR-Y3CB (regarding the ShotSpotter system in Chicago, "there is no district with a majority of White residents that has ShotSpotter wired up in their neighborhoods" while "ShotSpotter is deployed in every district that is above sixty five percent Latinx or Black." In Chicago, "[t]he upshot of this racially disparate ShotSpotter sensor deployment is that the negative consequences of ShotSpotter—including thousands of unsubstantiated police deployments—fall overwhelmingly on Chicago Black and Latinx residents.") [hereinafter State v. Williams Amicus Brief]; HELEN WEBLEY-BROWN ET AL., SHOTSPOTTER AND THE MISFIRES OF GUNSHOT TECHNOLOGY 12 (Surveillance Tech. Oversight Project 2017), https://perma.cc/JDD5-Y935 (reporting that ShotSpotter is deployed "almost exclusively" in communities of color in Cleveland, Atlanta, Kansas City, and New York City).

occurrence of false alerts from vehicle traffic, construction noises, and many other sources of noise in urban environments—all of which affect ShotSpotter's method[39]—the inevitable result is tens of thousands of encounters between people of color and police officers who have been prompted by ShotSpotter to believe that dangerous gun crime is afoot.[40] While people in majority White neighborhoods are generally not subjected to these ShotSpotter-initiated police encounters, people in communities of color in over 150 cities in the United States are exposed to this additional layer of policing as part of daily life.[41] Given this deployment strategy, it is no coincidence that Michael Williams and other people of color are the ones who are primarily harmed by ShotSpotter technology.

This Article discusses how ShotSpotter has flourished in the criminal justice system while avoiding meaningful oversight by the scientific community and rigorous vetting by the criminal justice system. Part I of this Article describes how SoundThinking's forensic method works in theory and examines how ShotSpotter is used by the police and prosecutors in the criminal justice system. Part II describes the scientifically-accepted process for the reliable development of forensic methods like SoundThinking's and the accumulation of the scientifically-meaningful data required for assessing forensic performance and calculating critical rates of error. Part III discusses SoundThinking's failure to comply with the accepted development processes, resulting in SoundThinking's promotion of performance and error rate claims that are not scientifically derived or justified. Part IV demonstrates that SoundThinking's failure to abide by the accepted development processes has resulted in the deployment of unreliable ShotSpotter systems in U.S. cities, with an unknown but significant error rate. Part V summarizes how the scientific and legal communities have failed to provide the oversight needed to vet ShotSpotter evidence and offers a plan for improved oversight. Finally, Part VI

---

[39] *See infra* notes 55-63 and 172-77 for a discussion of sources of ShotSpotter false alerts.

[40] State v. Williams Amicus Brief, *supra* note 38, at 16-17 ("The ShotSpotter system exposes individuals who live within its surveillance to a significant risk of harm and unjustified investigation at the hands of police officers responding to unreliable alerts. Each ShotSpotter alerts points police to a specific location and tells them someone is armed and has just fired their weapon. Individuals in the vicinity of an alert are immediately under suspicion by officers who are primed to believe that they are entering a dangerous situation. Meanwhile, residents in the vicinity will typically have no idea why police are descending on a particular spot—they won't know that ShotSpotter has interpreted some loud noise as a gunshot. This creates a highly volatile scenario, and [] can produce unwarranted investigatory stops, hostile encounters, and potentially dangerous intrusions on residents in the community.").

[41] *ShotSpotter Frequently Asked Questions*, *supra* note 3.

discusses the racial implications of the continued use of ShotSpotter technology in the absence of meaningful scientific and legal oversight.

I. SOUNDTHINKING'S BLACK-BOX FORENSIC METHOD AND ITS ROLE IN THE CRIMINAL JUSTICE SYSTEM

SoundThinking markets their ShotSpotter black-box gunshot detection method to local governments and police departments as both a real-time alert system for the detection of gunfire in urban communities as well as a source of reliable and admissible evidence for prosecutors to use in criminal cases.[42] SoundThinking claims that their ShotSpotter systems can quickly detect impulsive noises and subject those noises to a two-step forensic analysis—applying their algorithm followed by a human review—in order to classify noises as gunfire and determine its precise location. Importantly for the ShotSpotter method, the detection and both steps of the forensic analysis occur rapidly, so that real-time alerts—sent digitally from SoundThinking to police patrol officers—can trigger police responses and investigations. And when these ShotSpotter-initiated police responses result in arrests and criminal prosecutions, SoundThinking personnel offer expert testimony at criminal trials to the classification of detected noise events as gunfire as well as the calculated location of the noise events. In this way, criminal prosecutions are instigated by ShotSpotter alerts and prosecuted based on ShotSpotter claims.

A. SoundThinking's Black-Box Forensic Method

When police departments invest in ShotSpotter systems, they purchase a three-part forensic method designed to be a high-tech improvement over traditional 911 calls.[43] To accomplish this task, the ShotSpotter method involves (1) the collection noise events through acoustic sensors, (2) the use of a black-box algorithm to analyze noise events in an attempt to identify gunfire and calculate the precise location of the gunfire, and (3) a re-analysis by a human reviewer of the algorithm's determinations. And because SoundThinking seeks

---

[42] *Id.* (claiming that ShotSpotter is a "network of acoustic sensors that can detect, locate, and alert police to nearly all gunshot incidents. . . . [It] is used by police to: 1) be able to respond to a higher percentage of gunfire incidents; 2) improve response times to crime scenes to better aid victims and find witnesses; and 3) help police locate key evidence to identify and prosecute suspects.")

[43] *Save Lives and Find Critical Evidence with Proven Gunshot Detection,* SOUNDTHINKING https://perma.cc/X6FH-Z9XX (claiming that ShotSpotter is "like a digitized 911 call for gunshots that is faster and more accurate than our 50+ year old emergency call system").

to replace traditional 911 calls with their real-time alerts, the predominant goal of their forensic method is speed.[44] In fact, SoundThinking reports that it takes "less than 60 seconds" from the time a noise event occurs to the time that their forensic analysis is completed and a notification of gunfire reaches local police.[45]

The first step in SoundThinking's forensic method involves the detection of impulsive noises—sounds that are loudest at their inception and that dissipate quickly[46]—in targeted urban neighborhoods. To accomplish this step, SoundThinking deploys networks of microphones secured to light poles and similar structures that blanket selected neighborhoods. In the average ShotSpotter deployment, fifteen to twenty such microphones are installed per square mile.[47] These microphones instantly record audio clips of detected noises and wirelessly transmit the noise recordings and other related data to a central ShotSpotter hub for forensic analysis.[48]

The second step in SoundThinking's method is the forensic analysis of noise events using their proprietary ShotSpotter black-box algorithm. While most algorithms by definition share some level of opacity,[49] some are more opaque than others. Less sophisticated algorithms may be considered a "black-box" only because the developers have strategically chosen to not make the inner workings of their algorithms available for inspection "in order to maintain competitive advantage and/or to keep a few steps ahead of adversaries."[50] More complex types of black-box algorithms, such as "neural networks," are

---

[44] *Id.* (advertising that it takes less than sixty seconds from the occurrence of gunfire, through the ShotSpotter forensic analysis process, and to the issuance of a gunshot alert to local police officers).

[45] *ShotSpotter Frequently Asked Questions*, *supra* note 3.

[46] Andrew M. Willemsen & Mohan D. Rao, *Characterization of Sound Quality of Impulsive Sounds Using Loudness Based Metric*, 20TH INT'L CONG. ON ACOUSTICS, Aug. 2010, https://perma.cc/2PWQ-U6ZH.

[47] Test. of Paul Greene at 8, New York v. Simmons, No. 2016-0404 (Sup. Ct. N.Y. Monroe Cnty. Oct. 17, 2017) (on file with author).

[48] *ShotSpotter Frequently Asked Questions*, *supra* note 3.

[49] In discussions of algorithms and their function, the term opacity refers to "the lack of visibility of computational processes such that humans are not able to inspect the inner workings to ascertain for themselves how the results and conclusions were computed." Pragya Paudyal & B.L. William Wong, *Algorithm Opacity: Making Algorithmic Processes Transparent Through Abstract Hierarchy*, MIDDLESEX UNIV. LONDON, at 1 (2018), https://perma.cc/8BDH-ABDK; Jenna Burrell, *How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms*, BIG DATA & SOC'Y, Jan.-June 2016, at 1, https://perma.cc/LX9Z-948T ("[Algorithms] are opaque in the sense that if one is a recipient of the output of the algorithm . . . , rarely does one have any concrete sense of how or why a particular classification has been arrived at from inputs.").

[50] Burrell, *supra* note 49, at 3.

completely opaque in the sense that they "are no longer merely executing detailed pre-written instructions but are capable of arriving at dynamic solutions to problems based on patterns of data that humans may not even be able to perceive."[51] With these sophisticated algorithms, even the human developers who create the algorithms do not know how they operate, including which data the algorithms prioritize and what rules they follow to assess the data. While SoundThinking describes ShotSpotter as a "sophisticated machine algorithm[],"[52] no one outside of SoundThinking knows whether the opacity of their algorithm is a market strategy or a function of coding complexity.

Regardless of the explanation for the opacity of SoundThinking's black-box algorithm, their algorithm is tasked with several critical decisions. SoundThinking's algorithm receives sound recordings and other data from its deployed microphone networks and must instantly assess whether the recorded noises originated from gunfire or from other innocent sources.[53] For this classification task, the primary identifying characteristic of gunfire is its impulsivity—a sharp and loud onset of sound followed by a quick dissipation.[54] Complicating this classification task is the common occurrence of all manner of similar impulsive noises that can be mistaken for gunfire, from sources as varied as firecrackers,[55] car backfires,[56] loud mufflers,[57] jack hammers,[58] nail guns,[59]

---

[51] Bathaee, *supra* note 35, at 891.

[52] *ShotSpotter Frequently Asked Questions*, *supra* note 3.

[53] *Id.*

[54] Robert C. Maher & Tushar K. Routh, *Wideband Audio Recordings of Gunshots: Waveforms and Repeatability*, 141ST AUDIO ENG'G SOC'Y CONVENTION, Sept.-Oct. 2016, at 3, https://perma.cc/TRZ3-SJ3K (reporting on waveforms from the test firing of a rifle and handgun and explaining that these waveforms display "an abrupt spike as the muzzle blast wave arrives at the microphone" followed by a dissipation "as the muzzle blast energy dies away").

[55] Ihosvani Rodriguez, *Broward Sheriff Dropping Gunshot Detection System*, S. FLA. SUN-SENTINEL (Nov. 22, 2011, 12:00 AM), https://perma.cc/Z473-ZQ3M (reporting that Broward County Sheriff's Office quit using ShotSpotter because the system was "picking up noises such as firecrackers or a backfiring car and registering those sounds as gunfire. The sensors were also triggered by helicopters and the roar of downshifting trucks from nearby Interstate 95.").

[56] *Id.*

[57] Test. of Paul Greene, Trial Tr. Vol. 2, at 113, People v. Reed, No. 16015117 (Cal. Super. Ct. S.F. Cnty. July 6, 2017) (on file with author) [hereinafter July 6 Testimony in People v. Reed].

[58] Kara Grant, *ShotSpotter Sensors Send SDPD Officers to False Alarms More Often Than Advertised*, VOICE OF SAN DIEGO (Sept. 22, 2020), https://perma.cc/NRY8-JU3G.

[59] *Id.*

manual hammer strikes,[60] loud trucks,[61] helicopter noises,[62] and college campus noises.[63] This plethora of impulsive noises in urban environments poses a significant scientific challenge for ShotSpotter's algorithm during the classification process.[64]

In addition to this classification task, SoundThinking's black-box algorithm also attempts to decipher the locations of detected noise events in order to quickly provide police with the information needed to respond and search for suspects and victims.[65] To accomplish this task, SoundThinking has adopted an engineering concept called "multilateration."[66] Stripped down to its basics, multilateration involves using the time difference of arrival of soundwaves at multiple microphones to estimate the location of the origin of the noise event.[67] Assuming a more-or-less constant speed of sound,[68] the relative time that different microphones detect the same noise event allows for location estimations. For example, if soundwaves from a single noise event are detected by a microphone one second before they are detected by a second microphone, it is estimated that the origin of the sound event was about 340 meters closer to the first microphone.[69] Multilateration analysis uses these time differences across multiple microphones to estimate the location of the origin of the noise event.

---

[60] *Id.*

[61] July 6 Testimony in People v. Reed, *supra* note 57, at 113.

[62] Test. of Paul Greene, *supra* note 47, at 36.

[63] Kenneth C. Crowe II, *Troy Will Turn Off ShotSpotter*, TIMES UNION, https://perma.cc/Y4MS-SNSP (Oct. 30, 2012, 11:11 PM).

[64] *See infra* notes 165-77 discussing false alert instances encountered with ShotSpotter systems deployed in urban settings.

[65] *Save Lives*, *supra* note 43.

[66] Robert B. Calhoun et al., *Precision and Accuracy of Acoustic Gunshot Location in an Urban Environment*, ARXIV (Aug. 16, 2021, 11:54 PM UTC), https://perma.cc/HHT6-X7UT.

[67] ROBERT C. MAHER, PRINCIPLES OF FORENSIC AUDIO ANALYSIS 88-89 (2018) (ebook) ("Identifying a sound source location based upon simultaneous observations at two known positions uses a calculation procedure known as *multilateration*. Multilateration is based up on the *time difference of arrival* (TDOA) of a sound at two or more receive positions. . . . The principle of multilateration is that an impulsive sound produced by a source will propagate in all directions at the speed of sound, arriving at the receivers with a time delay corresponding to the source-to-receiver distance divided by the speed of sound." (emphasis in original)).

[68] AERONAUTICS RSCH. MISSION DIRECTORATE, SPEED OF SOUND (Nat'l Aeronautics & Space Admin. 2010), https://perma.cc/CF7V-JNRU (explaining that, while the speed of sound is sometimes estimated as 340 meters/second, various factors can affect the speed of sound, including air temperature, altitude, the density of the media through which soundwaves travel, and other factors).

[69] *See generally* Robert C. Maher & Ethan R. Hoerr, *Audio Forensic Gunshot Analysis and Multilateration*, 145TH AUDIO ENG'G SOC'Y CONVENTION, Oct. 2018, https://perma.cc/K6VH-W79K.

While using this multilateration technique in open farmland without obstacles to obstruct and redirect soundwaves is a less complex scientific challenge,[70] ShotSpotter's task of applying this technique in complex urban environments is a much more difficult scientific undertaking. The complexity of obstacles in urban environments—including dense buildings, raised highway networks, heavy vehicle traffic, and a multitude of other obstacles—causes soundwaves to bounce around the environment, get re-directed multiple times, and take highly irregular paths before eventually arriving at ShotSpotter microphones.[71] These irregular paths of soundwaves from their point of origin to detection microphones can result in significant uncertainty and error with ShotSpotter alerts because the relative difference in timing of the soundwaves becomes a function of random environmental variables rather than merely the speed of sound.[72]

For both ShotSpotter's classification and location tasks, the specific directions that SoundThinking has coded into their algorithm are unknown to anyone outside of the company. No one outside of SoundThinking knows how their algorithm has been trained to distinguish sounds originating from firecrackers or nail guns from the sounds of real gunfire. While the quality and

---

[70] Juan R. Aguilar, *Gunshot Detection Systems in Civilian Law Enforcement*, 63 J. AUDIO ENG'G SOC'Y 280, 286-87 (2015) (stating that when sensors are not within the line of sight of the source of sound waves, such as occurs in built environments, the performance of multilateration location estimates based on time difference of arrival calculations is reduced).

[71] Sylvain Cheinet et al., *Impulse Source Localization in an Urban Environment: Time Reversal Versus Time Matching*, 139 J. ACOUSTICAL SOC'Y AM. 128, 128 (2016) ("Many human activities take place in cities, and emit sound or noise. The localization of sound sources in urban environments is a topic of wide interest, with civilian needs as well as defense applications, e.g., for automotive safety, building engineering or shot localization. The acoustic sensing systems designed for such purposes need to adapt to the urban propagation physics between the source and the system. Among others, reflections and diffractions on buildings alter the times and angles of arrival of the acoustic waves. These effects deter the use of stand-alone acoustic antennas such as those used in open environments." (citations omitted)).

[72] Aguilar, *supra* note 70, at 286-87 ("The accuracy of shooter location estimates ranges from around 10 to 25 meters and is regarded as enough to identify shooter location in terms of street name and number. However, environmental issues affecting muzzle blast propagation in the outdoors imposes severe shortcomings on the accuracy of shooter location estimates. Most significant are the high sensitivity of gunshot detection algorithms to NLOS conditions, acoustic multipaths, background noise, and wind. For instance, at wind speeds of less than 3.5 ms^{-1} and at ranges of less than 250 m localization error is below 4% of the range to the shooter. Faster wind speeds of between 3.5 and 7 ms^{-1} would double the localization error. Moreover, multipath distortion tends to bias time delay estimates, which results in underestimations of the range to the shooter. In absence of line of sight to the shooter, the percentage of correct estimation of gunshot direction of arrival can be drastically reduced to less than 40%.").

quantity of data used to train algorithms like SoundThinking's means the difference between accurate performance and routine failure,[73] no one outside of SoundThinking has ever assessed ShotSpotter's training data for scientific sufficiency. Similarly, the modifications to basic multilateration coded into SoundThinking's algorithm to adapt the concept to complex urban environments are likewise a closely held secret. Consequently, neither the cities that purchase access to ShotSpotter systems nor the criminal justice system that relies on ShotSpotter evidence can know whether the training data or operating instructions of SoundThinking's algorithm are scientifically sound.

The third step in ShotSpotter's forensic method involves human examiners, described by SoundThinking as "[a]coustic experts."[74] The primary job of these examiners is to review and overrule algorithm classification judgments, switching the algorithm classification from a non-gunfire origin to gunfire, and vice versa.[75] But just like the secrecy shrouding SoundThinking's black-box algorithm, little is known about SoundThinking's human examiners. While forensic labs routinely disclose extensive documentation regarding examiner identities and qualifications including curriculum vitae, training documentation, proficiency testing histories, and other quality assurance documents,[76]

---

[73] See Steven Euijong Whang et al., *Data Collection and Quality Challenges in Deep Learning: A Data-Centric AI Perspective*, ARXIV (Dec. 26, 2022, 1:27 PM UTC), https://perma.cc/2EA2-6EXC (explaining that machine learning algorithm development requires that a majority of development resources go into selecting the data which will be used to train algorithms to accurately accomplish assigned tasks, reporting that "even the best machine learning algorithms cannot perform well" without good training data sets, and concluding that algorithms which are trained using "dirty, missing, or even poisoned data" can function maliciously); *see also* Jeffrey Dastin, *Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women*, REUTERS, https://perma.cc/RP72-QWA4 (Oct. 10, 2018, 4:04 PM) (reporting when Amazon, in an infamous example of algorithm training data failure, implemented an algorithm-based human resources system to expedite the process of reviewing and assessing the qualifications of candidates who applied for jobs with the company, and that after implementation, it was discovered that the algorithm routinely discriminated against female applicants because the data used to train the algorithm to assess qualifications was skewed toward prioritizing male applicants).

[74] *ShotSpotter Frequently Asked Questions*, *supra* note 352.

[75] Test. of Paul Green, Trial Tr. Vol. 1, at 16, 54, People v. Reed, No. 16015117 (Cal. Super. Ct. S.F. Cnty. July 5, 2017) (on file with author) [hereinafter July 5 Testimony in People v. Reed] (testifying that SoundThinking "incident review operators" listen to noise events, decide whether the noises are "likely to be from gunfire," and either report the noise event as suspect gunfire to police agencies or dismiss the noise events as non-gunfire; also testifying to a case example where the incident review operator "listened to the audio, looked at viewed audio wave for images, [and] made a judgment call to change the classification from possible gunshots to multiple gunshots").

[76] *See* NAT'L COMM'N ON FORENSIC SCI., RECOMMENDATION TO THE ATTORNEY GENERAL: NATIONAL CODE OF

SoundThinking regularly objects to these basic disclosures and seeks to maintain complete secrecy over examiner identities and qualifications.[77] So while very little is known about SoundThinking's forensic examiners, the limited amount of information about them in the public domain is not encouraging. Rather than requiring science-related bachelor's degrees like typical forensic examiners,[78] SoundThinking has sought to hire forensic examiners with "excellent customer service skills," a "can do attitude," and a "minimum of one year of professional experience, preferably in a call center."[79] And rather than providing the one to two years of formal forensic training offered to other forensic examiners,[80] SoundThinking offers their forensic examiners two to six weeks of on-the-job mentoring.[81]

PROFESSIONAL RESPONSIBILITY FOR FORENSIC SCIENCE AND FORENSIC MEDICINE SERVICE PROVIDERS (2016), https://perma.cc/DM3Q-2MW2 (stating that forensic science providers have ethical obligations to "[a]ccurately represent relevant education, training, experience, and areas of expertise," and to "communicate fully when requested with the parties through their investigators, attorneys, and experts"); *see also* NAT'L COMM'N ON FORENSIC SCI., RECOMMENDATION TO THE ATTORNEY GENERAL: TRANSPARENCY OF QUALITY MANAGEMENT SYSTEM DOCUMENTS (2016), https://perma.cc/C72W-Z2LD (stating that forensic laboratories should provide public access to all quality management system documentations, including proficiency testing, lab audits, and "curricula vitae for all analysts, scientists, and managers with positions of oversight over forensic testing").

[77] Third-Party Subpoena Recipient ShotSpotter, Inc.'s Motion to Quash Subpoenas *Duces Tecum* at 7-9, State v. Williams, No. 20cr0899601 (Ill. Cir. Ct. Cook Cnty. May 21, 2021) (on file with author).

[78] *See* SCI. WORKING GRP. ON FRICTION RIDGE ANALYSIS, STUDY & TECH., STANDARDS FOR MINIMUM QUALIFICATIONS AND TRAINING TO COMPETENCY FOR FRICTION RIDGE EXAMINER TRAINEES 1 (2012), https://perma.cc/54E3-V37Z (requiring fingerprint examiners to attain a bachelor's degree that "shall include science-related coursework"); *see also* SCI. WORKING GRP. FOR THE ANALYSIS OF SEIZED DRUGS, RECOMMENDATIONS 4 (2022), https://perma.cc/24CP-BZZS (requiring forensic drug examiners to obtain a science-related bachelor's degree and complete a documented training program with competency testing); *see also* NAT'L INST. OF JUST., U.S. DEP'T OF JUST., EDUCATION AND TRAINING IN FORENSIC SCIENCE: A GUIDE FOR FORENSIC SCIENCE LABORATORIES, EDUCATIONAL INSTITUTIONS, AND STUDENTS 7 (2004), https://perma.cc/7AX4-MAW8 (reporting on a consensus view in the forensic sciences regarding the critical content of undergraduate and graduate forensic science curriculum and asserting that "[a] model candidate for all forensic science practices . . . holds a baccalaureate degree (at a minimum) in the natural sciences").

[79] *Service Operations Center Specialist – Hiring All Shifts – FT/PT*, SHOTSPOTTER, https://perma.cc/928T-PHVU (Apr. 17, 2021, 10:40:53 GMT); *Incident Review Center Specialist – Hiring All Shifts – FT/PT*, SHOTSPOTTER, https://perma.cc/RJ8X-R78A (Feb. 5, 2023, 19:48:23 GMT).

[80] *See* SCI. WORKING GRP. ON FRICTION RIDGE ANALYSIS, STUDY & TECH., *supra* note 78 (requiring fingerprint examiners to undergo a training period of one to two years and pass competency testing at the conclusion of training to assess whether a sufficient level of competency has been attained); *see also* SCI. WORKING GRP. FOR FORENSIC DOCUMENT EXAMINATION, STANDARD FOR MINIMUM TRAINING REQUIREMENTS FOR FORENSIC DOCUMENT EXAMINERS (2013), https://perma.cc/FCU6-KEQK (requiring a minimum of twenty-four months of "full-time supervised training" for all forensic document examiners).

[81] July 6 Testimony in People v. Reed, *supra* note 57, at 155.

B. ShotSpotter's Use by Police and Prosecutors

Both police and prosecutors rely on ShotSpotter evidence when seeking to arrest and convict people of crimes. Police use ShotSpotter alerts as justification for stopping and searching people.[82] When conducting routine police patrols and not armed with search warrants, the U.S. Constitution only permits police to seize someone they encounter on the street when the police have individualized suspicion that the person is involved in criminal activity.[83] While the justification for such seizures is typically based on witness accounts of criminal activity or personal observations by police during patrol, police officers now rely on ShotSpotter for the justification to stop and search people.[84] In fact, police detain people based only on their proximity to ShotSpotter alerts.[85] In addition to the use of specific ShotSpotter alerts to pursue individual arrests, police departments collaborate with SoundThinking to use aggregated ShotSpotter alert data as the basis for strategizing larger police activities, such as the deployment of police resources to crime "hotspots."[86]

---

[82] OIG REPORT, *supra* note 14, at 18 (reporting the results of analysis of tens of thousands of ShotSpotter alerts in Chicago, documenting over 1,000 ShotSpotter-initiated stops during a 17-month period, and providing representative arrest narratives, such as "[arresting officers] were responding to a [ShotSpotter] of one round on the side of the building of [address]. [Arresting officers] observed [defendant] walking out from the side of the building at [address]. [Arresting officers] conducted an investigatory stop of the offender at above location.").

[83] Terry v. Ohio, 392 U.S. 1 (1968).

[84] OIG REPORT, *supra* note 14, at 19 (reporting that police officers in Chicago not only justify the detention of people based on their proximity to specific ShotSpotter alerts but also justify other detentions based on a person's proximity to previous ShotSpotter alerts, which "in the aggregate [] provide an additional rationale to initiate [a] stop or to conduct a pat down once a stop has been initiated").

[85] State v. Carter, 183 N.E.3d 611 (Ohio Ct. App. 2022) (finding the police detained and searched the defendant based solely on his proximity to a ShotSpotter alert); State v. Bellamy, No. A-2978-16T2, 2018 WL 2925724, at *1 (N.J. Super. Ct. App. Div. June 12, 2018) (finding the police stopped a man who was walking down the street because he was "the only person in the vicinity of the [ShotSpotter] reported gunshot."); State v. Martin, No. A-4026-18, 2021 WL 4592507, at *1 (N.J. Super. Ct. Law Div. Oct. 6, 2021) (finding the police stopped the vehicle that the defendant was in based on proximity to a ShotSpotter alert and a radio call about a "dark colored car with tinted windows"); Mitchell v. United States, 234 A.3d 1203, 1206 (D.C. 2020) (finding the police detained a bicyclist and searched him due to his presence in the area of a ShotSpotter alert); State v. Jobe, No. 2021AP712-CR, 2022 WL 1634777, at *1 (Wis. Ct. App. May 17, 2022) (finding the police initially stopped the defendant's vehicle due to the proximity of the defendant's vehicle to a ShotSpotter alert).

[86] Burke, *supra* note 4 (reporting that SoundThinking also offers an algorithm-based predictive policing method, which its claims can "forecast when and where crimes are likely to emerge and recommends specific patrols and tactics that can deter those events").

After police make ShotSpotter-initiated arrests, prosecutors pursue criminal convictions based on ShotSpotter evidence, using ShotSpotter alerts as the legal justification to move forward with criminal charges[87] as well as relying on SoundThinking employees to testify during criminal trials to the existence and location of gunfire.[88]

The prosecution of Christopher Carter for possessing illegal drugs provides one example of the way in which ShotSpotter alerts can result in police stops, arrests, and criminal convictions.[89] In this case, police in Dayton, Ohio were on routine patrol when they received a ShotSpotter alert—a notification that ShotSpotter had detected alleged gunfire in the vicinity of a particular address in the Dayton area. When police arrived in the area of the alert, they did not encounter any corroboration that gunfire had actually occurred—no fired casings, shooting victims, or witness accounts of gunfire. Nonetheless, police observed Mr. Carter walking in the area, briefly questioned him about why he was in the area, and then detained and searched him. While the police search did not uncover a weapon or any evidence of gunfire, police recovered methamphetamine from Mr. Carter's shorts and arrested him. Even though the absence of evidence that Mr. Carter was engaged in criminality at the time that police seized him rendered the police action and his resulting prosecution unconstitutional,[90] Mr. Carter was nonetheless successfully prosecuted for possession of methamphetamine and was eventually sentenced to prison.

---

[87] State v. Carter, 183 N.E.3d 611, 612-13 (Ohio Ct. App. 2022) (finding that prosecutors relied on ShotSpotter evidence alone as the justification for denying a motion to suppress evidence based on an unconstitutional seizure and search); United States v. Carter, Crim. No. 20-05 (JDB), 2020 WL 3893023, at *5 (D.D.C. July 10, 2020) (finding that prosecutors sought to justify a stop and seizure of the defendant due to "gunshots identified in the area just a minute before [police arrival] via the ShotSpotter system, combined with the fact that the neighborhood was a high-crime area"); United States v. Vallo, 608 F. Supp. 3d 1071, 1079 (D.N.M. 2022) (holding that, in a prosecution for possession of a firearm, the prosecution unsuccessfully sought to rebut the defendant's claim of an unconstitutional stop by asserting that the officers were justified in stopping and searching the defendant because of his presence in the area of a ShotSpotter alert).

[88] *SoundThinking's™ Response to Associated Press Article*, SOUNDTHINKING (Aug. 26, 2021), https://perma.cc/YX72-6HV8 (stating that SoundThinking employees have testified in over 200 criminal cases).

[89] 183 N.E.3d at 619-21.

[90] *Carter*, 2020 WL 3893023, at *6 (holding that seizures by police based on ShotSpotter alerts are unconstitutional because "[a]ttaching individualized suspicion to every person out and about in a residential area—or even to the first person sighted—merely because there were shots reported nearby, would incriminate 'a very large category of presumably innocent people'"); *In re* D.L., 147 N.E.3d 114, 119 (Ill. App. Ct. 2017) (finding no reasonable suspicion to conduct a *Terry* stop even though officers received multiple calls of shots fired

II. MEANINGFUL TESTING, VALIDATION, AND VERIFICATION OF RELIABLE FORENSIC TECH

Because "[f]orensic data, results, interpretations, and conclusions have life-changing consequences for individuals and society," the scientific community has developed a rigorous scheme to maximize the chance that forensic tech development occurs in a reliable way and to quantify the accuracy of methods like SoundThinking's at the end of the development process.[91] One central aspect of this scheme is the expectation that forensic methods like SoundThinking's undergo multiple levels of validation and error testing to establish that forensic conclusions are "generated through reliable methods and practices built upon valid core scientific principles and methodology."[92] Only through compliance with this rigorous process can a forensic method developer like SoundThinking generate the empirical data necessary to offer meaningful performance claims and to estimate critical rates of error.[93]

A. Validation Testing of New Forensic Methods

Regardless of whether a forensic method employs a black-box algorithm, validation testing of all new forensic methods is needed to document method performance, identify instances in which methods tend to fail, and provide empirical data for the estimation of meaningful error rates.[94] Generally,

---

from a particular intersection, first observed the Defendant walking quickly away from the scene just a few houses away, and then saw Defendant flee upon noticing the police).

[91] NAT'L COMM'N ON FORENSIC SCI., VIEWS OF THE COMMISSION: VALIDATION OF FORENSIC SCIENCE METHODOLOGY 1 (2016), https://perma.cc/S8CC-G6RR.

[92] *Id.*

[93] For a forensic method like SoundThinking's, there are several important rates of error and performance metrics that can be empirically estimated. For ShotSpotter's classification task of determining whether a noise originated from gunfire, important performance metrics include the method's false positive and negative rates, as well as the repeatability and reproducibility of examiner classification decisions. Separately, empirical testing can provide uncertainty estimations for ShotSpotter's additional task of location decisions. *See* PRESIDENT'S COUNCIL OF ADVISORS ON SCI. & TECH., *supra* note 34, at 47 ("For a metrological method to be scientifically valid and reliable, the procedures that comprise it must be shown, based on empirical studies, to be *repeatable*, *reproducible*, and *accurate*," where repeatability is defined as the rate at which "an examiner obtains the same result, when analyzing samples from the same source"; reproducibility is defined as the rate at which "different examiners obtain the same result, when analyzing samples from the same samples"; and accuracy is defined as the rate at which "an examiner obtains correct results both for samples from the same source (true positives) and for samples from different sources (true negatives)." (emphasis in original)).

[94] AM. ACAD. OF FORENSIC SCI. STANDARDS BD., *Foreword* to STANDARD FOR VALIDATION OF PROBABILISTIC GENOTYPING SYSTEMS (2020), https://perma.cc/NN8G-G6VD (stating that "[t]he validation of

validation testing involves "the acquisition of test data and determination of conditions and limitations of a new methodology."[95] For such testing to provide meaningful assessments of performance and valid error estimations, robust validation testing should satisfy three important criteria: the testing must (1) be accomplished under controlled conditions with known testing samples,[96] (2) simulate important variables that are likely to impact method performance during real-world deployment,[97] and (3) result in empirical data from which

computer software systems used for the probabilistic evaluation and interpretation of genetic information from forensic casework is a critical component of the validation process any caseworking laboratory using such software undergoes. Validations of such systems provide the study results and conclusions necessary for customers of forensic science service providers to have confidence in the evidence provided"); Kevin A. Schug, *Forensics, Lawyers, and Method Validation—Surprising Knowledge Gaps*, LCGC BLOG (July 23, 2015), https://perma.cc/SL6Z-DSLT (stating that the "failure to appropriately validate and document a method makes it impossible to prove the validity of the scientific test performed by that method" and asserting that "such a result would be scientifically unacceptable").

[95] HUM. FACTORS COMM., ORG. OF SCI. AREA COMMS. FOR FORENSIC SCI., HUMAN FACTORS IN VALIDATION AND PERFORMANCE TESTING OF FORENSIC SCIENCE 6 (2020), https://perma.cc/R7FW-5ZR3 [hereinafter OSAC REPORT].

[96] Jonathan J. Koehler, *How Trial Judges Should Think of Forensic Science Evidence* 102 JUDICATURE 28, 34 (2018) ("The most important indicator of the reliability of a forensic method is the rate at which trained examiners who use that method err: the lower the error rate, the greater the reliability of the method. Of course, in an actual case in which an unknown print or marking is compared to one or more knowns, ground truth is absent. In such cases, we cannot be sure whether a correct result is achieved because there is no independent way to verify the accuracy of the examiner's conclusion. But in a properly designed test in which prints or markings are produced from recorded knowns, ground truth is available, and an examiner's error rate . . . may be computed."); AM. ACAD. OF FORENSIC SCI. STANDARDS BD., STANDARD FOR VALIDATION STUDIES OF DNA MIXTURES, AND DEVELOPMENT OF VERIFICATION OF A LABORATORY'S MIXTURE INTERPRETATION PROTOCOLS 2 (2018), https://perma.cc/5FYK-DR84 [hereinafter AAFS LABORATORY DNA STANDARDS] (requiring that DNA validation testing be performed on samples which are "constructed from extracted DNA samples of known origin"); *see also* AM. ACAD. OF FORENSIC SCI. STANDARDS BD., STANDARD PRACTICES FOR METHOD VALIDATION IN FORENSIC TOXICOLOGY 4 (2019), https://perma.cc/6MM7-P7P3 [hereinafter AAFS TOXICOLOGY VALIDATION] (stating that toxicology methods must undergo validation using standard reference samples from known sources).

[97] OSAC REPORT, *supra* note 95, at 11 ("A key issue in validation is whether the test specimens adequately represent the range and difficultly of the items encountered in ordinary casework. If the study is designed to test the accuracy of a method for casework in general, then the samples should represent the full range and distribution of types and difficulty normally seen in casework."); *see also* Geoffrey Stewart Morrison et al., *Vacuous Standards—Subversion of the OSAC Standards-Development Process*, 30 FORENSIC SCI. INT'L: SYNERGY 206, 207 (2020), https://perma.cc/476T-LRWW (stating that method validation must occur "using data that reflect anticipated casework conditions"); SCI. WORKING GRP. ON DNA ANALYSIS METHODS, THE GUIDANCE DOCUMENT FOR THE FBI QUALITY ASSURANCE STANDARDS FOR FORENSIC DNA TESTING AND DNA DATABASING LABORATORIES 33-34, 35 (2020), https://perma.cc/MW5F-NQVP (requiring validation testing of complex mixture samples, samples of varying contributor ratios, and samples with varying template amounts and stating that "[m]ock samples should be reflective of the type and quality expected to be encountered in casework (e.g., various substrates, various stain concentrations)").

important error rates, including false positive and false negative rates, can be estimated.[98] Because this validation process is the only way a forensic method developer like SoundThinking can "determine the accuracy and limitations of the testing and interpretation parameters,"[99] the broader scientific community[100] and the more narrow forensic community[101] both consider robust validation testing to be a bright-line requirement for the reliable development and implementation of forensic methods. In fact, forensic laboratories cannot attain accreditation without providing proof that each of their forensic methods has undergone sufficient validation testing.[102]

New forensic methods should undergo such validation testing during at least two critical points of their development and deployment process.[103] During the initial development phases, the controlled testing required to assess

---

[98] OSAC REPORT, *supra* note 95, at 17 ("[T]here are two kinds of errors that the practitioner might make [when using a forensic categorical reporting scheme]: reporting an inclusion (i.e., that two items have the same source) when they in fact have different sources (a false inclusion); and reporting an exclusion when the items in fact have the same source (a false exclusion). Both kinds of errors (false inclusions and false exclusions) should be reported when presenting the results of a validation study.").

[99] *See* AM. ACAD. OF FORENSIC SCI. STANDARDS BD., STANDARD FOR THE DEVELOPMENT OF INTERNAL VALIDATION OF FORENSIC SEROLOGICAL METHODS 2 (2020), https://perma.cc/839Z-QPX7 (stating that separate internal validation testing must follow developmental validation in order to demonstrate that "the established protocols for the technical steps of the test and for data interpretation perform as expected in the laboratory").

[100] PRESIDENT'S COUNCIL OF ADVISORS ON SCI. & TECH., *supra* note 34, at 6 ("For forensic feature-comparison methods, establishing foundational validity based on empirical evidence is [] a *sine qua non*. Nothing can substitute for it.").

[101] OSAC REPORT, *supra* note 95, at 26 (stating that "[v]alidation is necessary in all scientific disciplines" due to the "consequences that may follow from a single forensic science analysis or comparison").

[102] *See* INT'L ORG. FOR STANDARDIZATION, ISO 17025: GENERAL REQUIREMENTS FOR THE COMPETENCE OF TESTING AND CALIBRATION LABORATORIES § 7.2.2.1 (2017) (stating that to achieve accreditation, "[t]he laboratory shall validate non-standard methods, laboratory-developed methods and standard methods used outside of their intended scope or otherwise modified").

[103] *See* Bruce Budowle et al., *Criteria for Validation of Methods in Microbial Forensics*, 74 APPLIED & ENV'T MICROBIOLOGY 5599, 5604 (2008) (stating that even after successful initial validation testing and deployment of a forensic method, subsequent validation testing must be undertaken whenever any significant changes to the forensic method occur); *see also* AAFS TOXICOLOGY VALIDATION, *supra* note 96, at 17 ("Modifications to a validated method shall be evaluated to confirm that the changes [to the method] do not have an adverse effect on the method's performance. The decision regarding which performance characteristics require additional validation shall be based on consideration of the specific parameters likely to be affected by the change(s)."); *see also* U.S. FOOD & DRUG ADMIN., ANALYTICAL PROCEDURES AND METHODS VALIDATION FOR DRUGS AND BIOLOGICS: GUIDANCE FOR INDUSTRY 10 (2015), https://perma.cc/43D8-UQS3 (describing revalidation procedures in the context of chemical analyses and stating that "[w]hen a change is made to an analytical procedure (e.g., a change in a piece of equipment or reagent or because of a change in manufacturing process or formulation), revalidation of all or part of the analytical procedure should be considered").

performance is called "*developmental validation*," which occurs "while the conditions and parameters are being worked out prior to the establishment of a defined assay, procedure[,] or product."[104] Developmental validation testing is used to assess whether the method development is faithfully implementing design plans and to provide baseline estimations of critical rates of error.[105] Developmental validation also provides developers with the testing data needed to understand the contexts in which a method performs best and the variables that can cause the method to generate unreliable results.[106]

A second round of validation testing in needed later in the process, once users like police departments and crime laboratories purchase and seek to deploy new forensic methods in real casework.[107] Such "*internal validation*" involves additional controlled testing designed to assess whether performance and error predictions generated during the development phase hold true when methods are deployed in real-world settings.[108] Because conditions and variables affecting method performance can differ from the development phase to real-world deployment, only after internal validation is completed can a forensic method developer offer scientifically defensible claims about performance and accuracy.

In the case of ShotSpotter, successful development and internal validation testing would need to involve live-fire testing which assesses the variables that are particular to the science of soundwave propagation.[109] These variables are well-known to the scientific community and include:

---

[104] OSAC REPORT, *supra* note 95, at 6.

[105] *Id.*

[106] Nicolas Hughes & Umit Karabiyik, *Towards Reliable Digital Forensics Investigations Through Measurement Science*, 2 WIREs FORENSIC SCI., no. 4, 2020, at 1 ("Validation, when conducted over the full range of conditions an analyst expects to encounter in evidence, provides objective, empirical data about the soundness and limitations of a particular forensic technique.").

[107] *See* AAFS LABORATORY DNA STANDARDS, *supra* note 96, at 1 (stating that separate internal validation testing must follow developmental validation in order to demonstrate that "the established protocols for the technical steps of the test and for data interpretation perform as expected in the laboratory").

[108] *Id.* ("[I]nternal validation [is] [t]he accumulation and evaluation of test data within the laboratory for developing the laboratory standard operating procedures and demonstrating that the established protocols for the technical steps of the test and for data interpretation perform as expected in the laboratory.").

[109] Aguilar, *supra* note 70, at 281 ("Outdoor propagation of muzzle blasts convey a number of related phenomena, including geometrical spreading, atmospheric absorption, wind and temperature gradients, turbulence, ground reflections, and acoustic multipath. Outdoor propagation phenomena could strongly affect the muzzle blast wave, particularly in the long distances encountered in gunshot detection scenarios.").

- Distance from gunshot to microphone;[110]
- Angle of gunshot to microphone;[111]
- Line-of-sight or non-line of sight status between gunshot and microphone;[112]
- Density of the microphone network;[113]
- Level of environmental noise;[114]
- The make and model of the gun, along with the type of ammunition discharged;[115]

---

[110] RYAN LILIEN, DEVELOPMENT OF COMPUTATIONAL METHODS FOR THE AUDIO ANALYSIS OF GUNSHOTS 1 (Off. of Just. Programs' Nat'l Crim. Just. Reference Serv. 2019), https://perma.cc/P5CU-E5ZZ ("If the blast is close to the recording device, the volume of the blast may overwhelm the recorder resulting in saturation and spectral information loss."); Steven D. Beck et al., *Variations in Recorded Acoustic Gunshot Waveforms Generated by Small Firearms*, J. ACOUSTICAL SOC'Y AM. 1748, 1754 fig.6, 1755 fig.7 (2011) (showing variations in gunshot waveforms due to distance from gunshot to microphone).

[111] LILIEN, *supra* note 110, at 2 ("[S]econdary factors effect the recorded audio. . . . [T]he muzzle blast is highly directional, dependent on the azimuth angle formed between the muzzle direction and the recording device.); Beck et al., *supra* note 110, at 1749 fig.1, 1755 fig.8 (showing variations in gunshot waveforms due to azimuth angle between gunshot and microphone); MAHER, *supra* note 67, at 118 fig.9.9 (documenting the results of live-fire testing and showing that a sensor directly behind the shooter position did not detect a soundwave while sensors in front and beside the shooter position detected soundwaves)

[112] Aguilar, *supra* note 70, at 284, 286-87 ("Once a possible gunshot has been detected, the next step is to discriminate if the signal corresponds to an actual gunshot or if it constitutes a different type of high-amplitude impulse sound. This could be a quite problematic task and very susceptible to environmental issues such as background noise, acoustic multipath, and NLOS condition. . . . [E]nvironmental issues affecting muzzle blast propagation in the outdoors imposes severe shortcomings on the accuracy of shooter location estimates. Most significant are the high sensitivity of gunshot detection algorithms to NLOS conditions, acoustic multipaths, background noise, and wind.").

[113] Calhoun et al., *supra* note 66, at 2, 7 (stating that the ShotSpotter detection method "works best when the sensor density is high enough to ensure gunshots are detected on at least two more sensors than required for multilateration" and also stating that higher sensor densities are needed in areas "with a high density of structures, high background noise (traffic, rapid transit systems), difficult wind conditions, or unfamiliar environments").

[114] Izabela L. Freire & José A. Apolinário Jr., *Gunshot Detection in Noisy Environments*, 7TH INT'L TELECOMM. SYMP., Mar. 2010, at 3 tbl.3, https://perma.cc/K6KS-QTEW (showing that noisier environments create more false positive errors for gunshot detection systems); William Renda & Charlie H. Zhang, *Comparative Analysis of Firearm Detection Recorded by Gunshot Detection Technology and Calls for Service in Louisville, Kentucky*, 8 INT. J. GEO-INF. 275, 276 (2019) ("Heavily noisy environments, such as real-world urban settings, have been shown to affect [gunshot detection] effectiveness where up to 9% of actual gunfire is not detected and approximately 25% of non-gunfire events with a similar acoustic signature, i.e., balloon popping and hand clapping, were falsely identified as gunfire.").

[115] LILIEN, *supra* note 110, at 2 ("The recording is also influenced by the firearm make/model, caliber, and ammunition type. Each device has a frequency response that describes how efficiently the device captures sound at different frequencies."); Robert C. Maher & Steven R. Shaw, *Directional Aspects of Forensic Gunshot Recordings*, ANTENNAS & ELECTROMAGNETIC SYS.

- Density and variation of urban structures and landscape;[116]
- Time of day;[117]
- Weather conditions such as temperature, rain, thunder, and wind.[118]

Because each of these variables can affect the performance and accuracy of noise detection systems like ShotSpotter, assessing and documenting their impacts on performance and rates of error during validation is critical.

B. Black-Box Algorithms Require Additional Safeguards

While this two-step developmental and internal validation process suffices for traditional forensic methods, an additional process of verification and validation (V&V)[119] is needed for forensic methods like SoundThinking's that employ black-box algorithms.[120] In order to separately assess whether the

---

39TH INT'L CONF., June 2010, at 5, https://perma.cc/AG6L-L6E7 ("The on-axis waveforms for the ten firearms are shown in Figure 12 with the same amplitude scale for each waveform. . . . For visual examination, the gunshot waveform features show noticeably different and distinct waveshapes.").

[116] Aguilar, *supra* note 70, at 281 ("The presence of surrounding buildings that compose the urban landscape also affects the muzzle blast propagation phenomena as it introduces multipath distortion, acoustic diffraction, and non-line-of-sight (NLOS) conditions between shooter and sensor locations."); Tony F. W. Embleton, *Tutorial on Sound Propagation Outdoors*, 100 J. ACOUSTICAL SOC'Y AM. 31, 35 (1996) ("[A]sphalt and grass-covered grounds have different effects on those parts of a sound field that propagate near to the ground surface. It is necessary to be able to categorize these and other commonly occurring ground surfaces such as concrete, snow, or earth in order to be able to predict their effects.").

[117] Renda & Zhang, *supra* note 114, at 276 (reporting that "[t]he accuracy and sensitivity of [gunshot detection devices] to detect actual gunfire has been shown to vary spatially and temporally, with better performance at nighttime and with increased density of sensors").

[118] LILIEN, *supra* note 110, at 2 ("Finally, the audio is susceptible to environmental conditions (temperature, humidity, wind) and scene geometry (absorption, reflection, focusing)"); Tsiatsis E. Nikolaos, *Recording and Calculating Gunshot Sound—Change of the Volume in Reference to the Distance*, 1203 AM. INST. PHYSICIANS CONF. PROC. 846, 851 (2010) ("There are several different factors which influence the volume of the gunshot intensity, [such] as the following: [] The length of the gun barrel (the shorter the barrel, the louder the sound), [] The powder of the ammunition that is used for the fire, [] The speed/direction of the wind."); Embleton, *supra* note 116, at 31 ("Wind and temperature gradients in the atmosphere cause refraction which can either increase or decrease sound pressure levels significantly.").

[119] In general, software validation and verification are the "methods and technologies that provide confidence in system software." *See* Yinghua Guo et al., *Validation and Verification of Computer Forensic Software Tools—Searching Function*, 6 DIGIT. INVESTIGATION S12, S13 (2009).

[120] Marc Canellas, *Defending IEEE Software Standards in Federal Court*, 54 COMPUT. 14, 17 (2021) ("Scientists, engineers, and IEEE 1012 have long demanded that safety-critical software and hardware be the right systems built the right way, and the law should demand this, too. . . . Sponsored by the IEEE Computer Society, IEEE 1012 is a universally applicable and broadly accepted process for ensuring that the right product is correctly built for its

algorithm employed in a forensic method is fit for its designated purpose,[121] the scientific community has promulgated clear and accessible V&V standards for assessing computer code during development. One of the most prominent standards in the scientific community,[122] IEEE 1012 Standard for System, Software, and Hardware Verification and Validation[123] ("IEEE 1012"), requires extensive code assessments designed to answer critical questions, including "can computer predictions be used as a reliable bases for crucial decisions" and "[w]hat confidence can be assigned to a computer prediction of a complex event."[124] The level of code scrutiny required by IEEE 1012 depends on the seriousness of harm to society should the computer code fail. Because the use of software "in criminal court can result in catastrophic failures through false imprisonment and the deprivation of people's rights," the most rigorous level of V&V is required for algorithms like SoundThinking's.[125] The auditing process associated with algorithm verification has been embraced by the aviation industry, industrial control industry, and some financial sectors.[126]

---

intended use."); Guo et al., *supra* note 119, at S12 ("As today's [electronic evidence] investigations heavily rely on automated software tools, the reliability of investigation outcomes is predominantly determined by the validity and correctness of such tools and their application process. Therefore, an insistent demand has been raised by law enforcement and other agencies to validate and verify [electronic evidence] tools to assure the reliability of digital evidence.").

[121] Dolores R. Wallace & Roger U. Fujii, *Software Verification and Validation: An Overview*, 6 IEEE SOFTWARE 10, 10 (1989) (describing software verification and validation as a structured approach to analyzing software during development "to determine that it performs its intended functions correctly, to ensure that it performs no unintended functions, and to measure its quality and reliability").

[122] The Institute of Electrical and Electronics Engineers (IEEE) is the "largest organization of technology professionals in the world, representing more than 400,000 engineers, scientists, and allied professionals worldwide." IEEE-USA, Comment Letter on NIST Internal Report 8351-DRADT *DNA Mixture Interpretation: A NIST Scientific Foundation Review* (Nov. 18, 2021), https://perma.cc/54ZL-JZX3. In that capacity, the IEEE is "the leading developer of global technical standards used in power and energy, telecommunications, biomedical and healthcare, information technology, transportation, and information assurance products and services." *Id.* For information on the IEEE, see *id.*

[123] Canellas, *supra* note 120, at 17 ("IEEE 1012 was developed by the IEEE Standards Association (SA), a world-leading standard-setting organization (SSO) with its own reputation for developing reliable and fair regulations.")

[124] Ivo Babuska & J. Tinsley Oden, *Verification and Validation in Computational Engineering and Science: Basic Concepts*, 193 COMPUT. METHODS APPLIED MECHS. & ENG'G 4057, 4057 (2004).

[125] Canellas, *supra* note 120, at 17-18 (explaining that achievement of the level of IEEE 1012 V&V applicable to ShotSpotter would require that the technical staff who conduct the assessment not be members of the developing company, the managers who oversee the code assessment not have any formal affiliation with the developers, and the financing which would support the assessment process come from some source other than the developers).

[126] Joshua A. Kroll et al., *Accountable Algorithms*, 165 U. PA. L. REV. 633, 662 n.97 (2017)

Compliance with these V&V requirements is not a mere formality. Rather, after-the-fact audits of computer algorithms commonly used in the criminal justice system have uncovered serious algorithmic failure. One algorithm used in the criminal justice system to interpret DNA results and supply juries with evidence of guilt contained hidden coding errors that resulted in incorrect interpretations in at least sixty cases.[127] Coding flaws in a different DNA interpretation algorithm caused that algorithm to routinely "overestimate the likelihood of guilt."[128] And flaws in a breathalyzer algorithm's source code resulted in the dismissal of thousands of criminal cases in Massachusetts and New Jersey after questions were raised about the reliability of the breathalyzer results.[129] These examples and others[130] highlight the high stakes involved with the deployment of algorithms in the criminal justice system and the need for robust V&V.

C. Additional Error Analysis for Forensic Methods That Rely on Subjective Human Decision-Making

For forensic methods that involve the collaboration of algorithms and human examiners,[131] it makes little sense to comprehensively assess the accuracy and performance of the algorithm and ignore the role that the human

---

[127] Murray, *supra* note 24.

[128] State v. Pickett, 246 A.3d 279, 296 (N.J. Super. Ct. App. Div. 2021).

[129] *Id.*

[130] *See supra* notes 24-27 discussing problems with several algorithm-based methods used in the criminal justice system.

[131] Some forensic methods involve analysis and decision-making steps by both computer algorithms and humans. But the impacts of these AI-human collaborations are not well known yet in the scientific community. While initial research is beginning to quantify the subtle influences that algorithmic analyses and decisions have on the human examiners, the understanding of this dynamic is still incomplete. One example of AI-human forensic collaboration and the influences of algorithmic decisions involves the forensic fingerprint comparison process. The modern practice of forensic fingerprint comparisons often involves initial algorithmic searches of large fingerprint databased in order to generate lists of candidates who may be the source of crime scene prints. When forensic examiners first started utilizing the algorithm-generated candidate lists in order to assist in the forensic identification process, little was known about the subtle ways in which the algorithmic decisions influenced subsequent decisions by the human examiners. Initial research investigating the algorithmic influences has found that fingerprint examiners are unconsciously influenced by the algorithmic decisions to conduct less thorough forensic examinations of some candidates and can even be influenced by the algorithm candidate lists to commit more errors. *See* Itiel E. Dror et al., *The Impact of Human-Technology Cooperation and Distributed Cognition in Forensic Science: Biasing Effects of AFIS Contextual Information on Human Experts*, 57 J. FORENSIC SCI. 343, 350-51 (2011).

examiners play in the accuracy of final forensic decisions.[132] When human examiners use their subjective judgments[133] to second-guess algorithmic forensic decisions, additional controlled testing is needed to quantify the impact that examiners have on method performance and accuracy.[134] This controlled testing most often takes the form of large-scale error rate studies, involving many forensic examiners in individual forensic disciplines conducting thousands of forensic analyses on known samples.[135] Conducted by scientific organizations that are independent of forensic method developers, this testing provides critical estimations of error rates for human examiners, as well as other important metrics like repeatability and reproducibility.[136] Such human

---

[132] *Id.* at 343 (discussing how "[t]he increased use and reliance on technology have reached a level whereby humans and technology are more and more intertwined and collaborating with one another, creating distributed cognition." Thus, "[u]nderstanding each mode, both its potential and its limitations, is necessary to make optimal use of both the technological and the human elements in the collaboration. In other words, the success of human experts and technology working in such close collaborations depends on correctly distributing the work among them, taking advantage of the relative strength each has to offer, and avoiding their respective weakness and vulnerabilities."); Linda J. Skitka et al., *Does Automation Bias Decision-Making?*, 51 INT'L J. HUM.-COMPUT. STUD. 991 (1999) (reporting the results of experimentation showing that the introduction of computer aids to human decision making creates opportunities for humans to make different, and sometimes more, errors).

[133] PRESIDENT'S COUNCIL OF ADVISORS ON SCI. & TECH., *supra* note 34, at 49 ("Subjective methods require careful scrutiny, more generally, their heavy reliable on human judgment means that they are especially vulnerable to human error, inconsistency across examiners, and cognitive bias. In the forensic feature-comparison disciplines, cognitive bias includes the phenomena that, in certain settings, humans (1) may tend naturally to focus on similarities between samples and discount differences and (2) may also be influenced by extraneous information and external pressures about a case.").

[134] *Id.* ("Since the black box in the examiner's head cannot be examined directly for its foundational basis in science, the foundational validity of subjective methods can be established *only* through empirical studies of examiner's performance to determine whether they can provide accurate answers; such studies are referred to as 'black-box' studies. In black-box studies, many examiners are presented with many independent comparison problems—typically, involving 'questioned' samples and one or more 'known' samples—and asked to declare whether the questioned samples came from the same source as one of the known samples. The researchers then determine how often examiners reach erroneous conclusions." (emphasis in original) (citation omitted)).

[135] *Id.* at 46 ("For subjective feature-comparison methods, appropriately designed black-box studies are required, in which many examiners render decisions about many independent tests (typically, involving 'questioned' samples and one or more 'known samples) and the error rates are determined. Without appropriate estimates of accuracy, an examiner's statement that two samples are similar—or even indistinguishable—is scientifically meaningless: it has no probative value, and considerable potential for prejudicial impact. Nothing—not training, personal experience nor professional practices—can substitute for adequate empirical demonstration of accuracy.")

[136] Bradford T. Ulery et al., *Repeatability and Reproducibility of Decisions by Latent Fingerprint Examiners*, 7 PLOSONE, no. 3, Mar. 2012, at 1, https://perma.cc/2DU8-VJNL

error testing is so fundamental and expected in the forensic community that one group of influential scientists, the President's Council of Advisors on Science and Technology,[137] has commented that no evidence from any forensic discipline should be admissible in court prior to the publication of multiple such large-scale error studies.[138]

When such examiner performance data is published in peer-reviewed scientific journals, it provides critical insights into the reliability of forensic evidence offered in court. For instance, before the publication of the first large-scale error rate studies involving forensic fingerprint examiners, testifying examiners routinely overstated the value of fingerprint evidence by claiming that their method was objective and that there was no chance of error when examiners applied the method in casework.[139] These examiner claims were false,[140] but the data to prove falsity was not readily available until the publication of large-scale examiner performance studies. From just one such study involving the participation of 169 fingerprint examiners, claims of objectivity and certainty were refuted by empirical data that quantified the rate at which fingerprint examiners make certain types of errors and the frequency with which examiners disagree when examining the same forensic evidence.[141] Only after the publication of this empirical data did forensic fingerprint

(explaining that repeatability refers to intra-examiner agreement: whether one examiner consistently reaches the same decision when assessing the same evidence on multiple occasions, and also explaining that reproducibility refers to inter-examiner agreement: whether two or more examiners reach the same decision when assessing the same evidence).

137 *President's Council of Advisors on Science and Technology,* THE WHITE HOUSE, https://perma.cc/2DZF-LWNM (The President's Council of Advisors on Science and technology is "the sole body of advisors from outside the federal government charged with making science, technology, and innovation policy recommendations to the President and the White House" and is comprised of "distinguished individuals from industry, academia, and non-profit organizations with a range of perspectives and expertise.").

138 PRESIDENT'S COUNCIL OF ADVISORS ON SCI. &TECH., *supra* note 34, at 47.

139 Joseph B. Kadane & Jonathan J. Koehler, *Certainty and Uncertainty in Reporting Fingerprint Evidence*, 147 DAEDALUS 119, 120 (2018) ("Although the [fingerprint comparison] process is subjective, fingerprint examiners have historically claimed that their identifications are 100 percent certain, and that there is virtually no chance that an error has occurred.").

140 Ulery et al., *supra* note 136, at 6, 11 (reporting that fingerprint examiners in this study committed false positive errors at a rate of one in every 645 comparisons and further reporting that groups of examiners disagreed on the ultimate forensic conclusion 45% of the time when conducting comparisons on more challenging casework samples).

141 *See generally* Ulery et al., *supra* note 136.

examiners slowly start to modify their claims about the reliability of fingerprint evidence and provide more accurate testimony in court.[142]

III. SOUNDTHINKING'S FLAWED TESTING PROCESS AND THEIR UNRELIABLE PERFORMANCE AND ERROR CLAIMS

Despite having a scientific obligation to engage in the multi-step process described above for the development and implementation of a new forensic method, SoundThinking's approach to this process has been scientifically inadequate. Of the validation and testing processes describe above—developmental validation, internal validation, independent algorithm V&V, and large-scale human examiner error testing—SoundThinking has ignored all but the developmental validation step.[143] The lack of human error testing is of special concern considering the high level of disagreement between SoundThinking's forensic algorithm and their human examiners[144] and the fact that human examiners get the final word on whether to classify detected noises as gunfire events.[145] With regard to developmental validation, SoundThinking's efforts at accomplishing this step have been minimal. The result of this validation and testing failure is that SoundThinking possesses no legitimate scientific data upon which to base meaningful performance and accuracy claims about their ShotSpotter gunshot detection method. In the absence of real

---

[142] Kadane & Koehler, *supra* note 139, at 120 (reporting that after decades of overstated conclusions by fingerprint examiners, "federal agencies and forensic science professional organizations began working in earnest to, among other things, modify the ways in which forensic scientists present evidence in court. Simple and obvious reforms such as eliminating references to '100 percent certain' identifications and '0 percent risk of error' have already taken hold.")

[143] City of Chi., *Committee on Public Safety Hearing*, VIMEO, at 2:07:58-2:08:22 (Nov. 12, 2021), https://perma.cc/T5BL-YMPV (where ShotSpotter CEO Ralph Clark admitted that ShotSpotter does not conduct internal validation on deployed systems while claiming that such performance analysis is not necessary because ShotSpotter "[has] been around in business for a very long time" and "there's just a lot of experience out there that really speaks to our technical efficacy."); Letter from Warrington Parker, Attorney, to author (June 1, 2022) (on file with author) (where the defense requested the production via subpoena of "all validation and verification of the computer code involved in ShotSpotter's gunshot detection/location/timing system" and received neither a written V&V plan nor any indication that ShotSpotter has sought to comply with the independent V&V processes applicable to its forensic method).

[144] Letter from Gary Bunyard to Chairman Taliaferro and Members of the Joint Committee on Public Safety and Health and Human Relations (Feb. 3, 2022) (on file with author) (based on eleven months of ShotSpotter alert data from Chicago during 2021, ShotSpotter examiners rejected 648,332 (94%) algorithm gunfire classifications and affirmed only 43,060 (6%) algorithm gunfire classifications).

[145] July 5 Testimony in People v. Reed, *supra* note 75, at 16, 49, 54.

accuracy data, SoundThinking promotes accuracy claims that are based on unscientific sources, such as survey responses and anecdotes. In doing so, they mislead their police department clients and criminal justice end users about the performance of their forensic method.

A. SoundThinking's Flawed Testing Process

While ignoring their scientific obligation to conduct internal validation testing and independent algorithm V&V, SoundThinking has attempted limited live-fire developmental validation testing.[146] But SoundThinking's developmental validation efforts represent only a small and flawed first step in the developmental validation process.[147] SoundThinking's development testing attempt involved the type of live-fire testing required for validation, and their resulting study documented the performance of their algorithm in detecting and locating gunfire noises.[148] But the flaws and limitations with this study

---

[146] *See* Calhoun et al., *supra* note 66, at 9-10.

[147] Although there are two additional documented live-fire validation attempts of ShotSpotter, neither represents meaningful validation of its current forensic method due to the age of the studies, the significant changes to ShotSpotter's forensic method since those validation attempts, and the flaws and limitations with these two studies. The oldest of these studies involved live-fire testing in 1997. LORRAINE G. MAZEROLLE ET AL., FIELD EVALUATION OF THE SHOTSPOTTER GUNSHOT LOCATION SYSTEM: FINAL REPORT OF THE REDWOOD CITY FIELD TRIAL (1999), https://perma.cc/W3LE-7LXA. The other study involved similar testing in 2006. NAT'L L. ENF'T & CORR. TECH. CTR., CUSTOMER ACCEPTANCE TEST REPORT: WIRELESS GUNSHOT LOCATION SYSTEM WITH ALVARION RADIOS (2006) (on file with author). The current ShotSpotter forensic method has changed significantly since 1997 and 2006. For instance, forensic analysis of noise events, which used to be conducted by local police officials in individual police departments, was transferred in 2011 to centralized Incident Review Centers staffed by SoundThinking employees. July 5 Testimony in People v. Reed, *supra* note 75, at 20-21. Additionally, both the hardware and software that ShotSpotter uses for critical steps of pulse detection and signal processing has changed over time. Finally, it is likely that the algorithms SoundThinking uses for classification and location estimation and its training data are different now than the ones used in 1997 and 2006. Also, both of these studies failed to assess and estimate the most important type of error (false positive errors) encountered with ShotSpotter systems and failed to document the impacts of important variables on false negative error rates. *See* Budowle et al., *supra* note 103, at 5604 (stating that the predictions and performance claims resulting from the validation process are only valid as long as there are no significant changes to the forensic method or the environment in which it operates: "The validation process is not a one-time event for a method. It must be considered dynamic in order to assess periodically the impact of new knowledge and findings to assess material modifications made to existing methods and procedures. Indeed, monitoring and reassessment are tools to ensure that even previously validated processes remain valid if the parameters under which the process is carried out are altered."); *see also* U.S. FOOD & DRUG ADMIN., *supra* note 103, at 9 (stating that scientific methods need to be monitored during their life cycle and recommending that methods be "reevaluated, revalidated, or amended, as appropriate" when operational conditions change).

[148] Calhoun et al., *supra* note 66, at 9-15.

render it insufficient to validate SoundThinking's forensic method and measure ShotSpotter's true performance. Most importantly, this study did not attempt to measure the most critical type of error generated by the ShotSpotter gunshot detection method: false positive errors, where ShotSpotter falsely reports innocent noises as gunfire.[149] Additionally, while this testing resulted in the reporting of data about ShotSpotter's false negative rate, the reported error rate is of limited scientific value because this testing failed to assess and document how ShotSpotter's ability to accurately detect real gunfire is impacted by common variables encountered in the real world—the distance from sound sources to ShotSpotter microphones, the density of urban structures, the line-of-sight status between sound source and ShotSpotter microphones, the level of environmental noise, and other variables.[150] Ultimately, the conditions under which this testing occurred are not representative of many of the conditions that ShotSpotter encounters on a daily basis in the real world.[151] Instead, the company tested ShotSpotter's

---

[149] Murrie et al., *supra* note 30, at 3 ("A false positive error in forensic science conclusions typically results in criminal charges against an innocent individual whereas false negative errors result in guilty individuals avoiding legal charges. . .analysts reported that they, their workplace, and their discipline prefer to minimize the risk of false positive errors and thus tolerate a greater risk of false negative errors.").

[150] *See supra* notes 110-18, discussing the variables that should be assessed during ShotSpotter validation testing, including line-of-sight status, sensor density, distance from noise source to sensor, angle of the noise event in relationship to sensors, level of environmental noise, and the density of urban structures in the vicinity of the noise event.

[151] Contrast ShotSpotter's validation efforts to those involved with a similar gunshot detection system called SECURES. *See* MICHAEL LITCH & GEORGE A. ORRISON, IV, DRAFT TECHNICAL REPORT FOR SECURES DEMONSTRATION IN HAMPTON AND NEWPORT NEWS, VIRGINIA (2011), https://perma.cc/VKT6-LZ4P. For the SECURES live-fire testing, the planning and execution of the testing was handled by agencies (U.S. Department of Justice and the Center for Society Law and Justice) independent of the method developer. *Id.* at v. Additionally, this validation testing sought to empirically quantify both the false negative rate and the false positive rate—as well as other important performance metrics—of the gunshot detection system. *Id.* at 11. Importantly, the study leaders recognized that meaningful rates of error could only come from controlled validation data and not operational data, conceding that "[d]ata from the operational test period—dispatching system records and filed officer reports—can be used to answer a variety of questions about the usefulness of automated gunshot detection for law enforcement" but "[s]uch data cannot, however, be used for assessing the accuracy of the system for detecting or localizing actual gunshots." *Id.* at 20. For this reason, this validation study involved the intentional activation of sources of confounding impulsive noises, including firecrackers and other fireworks. *See, e.g.*, *id.* at 20, tbl.2. Through the analysis of this additional empirical performance data, the study leaders were able to offer empirically-based judgments about the susceptibility of the system to one source of false positive errors. *Id.* at 26. Finally, these study leaders recognized that the performance metrics reported for this validation process only provided important insights into method performance for the two cities involved in the testing, stating that "[a] law enforcement

performance under "best case scenarios" for successful performance: in a setting with an unusually high density of ShotSpotter microphones, in open-air parking lots and playing fields—locations that lack large multi-unit buildings typical of many urban neighborhoods—and during a quiet autumn night.[152] Taken together, these flaws with SoundThinking's attempt at developmental validation testing should result in very little credence given to SoundThinking's performance claims stemming from this testing.[153]

B. SoundThinking's Unreliable Performance and Error Claims

In the absence of scientifically-sufficient validation and error testing as the basis for ShotSpotter's performance and accuracy claims, SoundThinking has promoted accuracy claims that have no meaningful scientific bases. As support for their claims that ShotSpotter is accurate and reliable, SoundThinking routinely references two documents that do not demonstrate legitimate scientific testing and only provide unscientific and speculative bases for their claims. SoundThinking relies on the first of these documents—entitled "Independent Audit of the ShotSpotter Accuracy"[154]—as the primary basis for their accuracy claims. The authors of this document state that they accumulated "complete and accurate" error data from ShotSpotter, which they "validated" to generate "robust" results.[155] Based on this process, the authors reported an overall accuracy rate of 97.59% and a separate false positive rate of 0.41%.[156] The second document, entitled "Gunshot Location System Efficacy

agency must determine the parameters of these tradeoffs [between detection of real gunshots and the reporting of false alerts] themselves through careful testing in each of its coverage areas." *Id.* at 27.

[152] *See* Renda & Zhang, *supra* note 114, at 276 ("The accuracy and sensitivity of GDT to detect actual gunfire has been shown to vary spatially and temporally, with better performance at nighttime and with increased density of sensors."); *see also* Maher & Routh, *supra* note 54, at 2 ("The muzzle blast acoustical characteristics depend upon the type and size of the firearm, the characteristics of the ammunition, the direction with respect to the barrel axis, the presence of acoustical reflections from nearby surfaces, and diffraction from nearby obstacles."); Embleton, *supra* note 116, at 31.

[153] *See* Jennifer L. Mnookin et al., *The Need for a Research Culture in the Forensic Sciences*, 58 UCLA L. REV. 725, 742 (2011) ("Research that is deeply methodologically flawed should be given no credence. Moreover, research that is methodologically sound should not be touted as offering support for propositions that extend beyond the reach of the research design. In short, the extent of sound empirical support for claims should guide practices in the laboratory, conclusions in reports, and testimony in the courtroom.").

[154] EDGEWORTH ANALYTICS, INDEPENDENT AUDIT OF THE SHOTSPOTTER ACCURACY (2022) https://perma.cc/4TNR-UWL7.

[155] *Id.* at 2-3.

[156] *Id.* at 1.

Study," states that "ShotSpotter's accuracy, of both geographic location of an incident and of shots, is its best attribute."[157] These authors further claim that false negative errors by ShotSpotter are "very rare" and that false positive errors, though they can occur, "do not diminish and indeed are orthogonal to the general efficacy of the ShotSpotter product."[158] SoundThinking representatives reference these accuracy claims when seeking contracts with local governments to purchase and install their forensic systems.[159]

Despite SoundThinking's reliance on these two documents as the bases for performance claims, neither document contains scientifically-valid empirical evidence of true method performance. Rather, the authors of both documents base their claims on feedback and anecdotes from police agencies and not robust scientific testing. The first study's claims of a 97.59% accuracy rate and a 0.41% false positive rate are based on customer feedback: the authors tabulated "potential errors *identified by clients* for investigation and *ShotSpotter's conclusions* regarding those potential errors."[160] In other words, for a ShotSpotter alert to be categorized as an error, the underlying incident had to have been known to police officers, reported through the police chain of communication and passed onto SoundThinking personnel, and not otherwise determined after the fact by SoundThinking to have been accurate. Because police are unaware of most instances of gunfire,[161] they cannot report even a fraction of those instances missed by ShotSpotter. And because police departments lack the time and will to document and report the many known potential ShotSpotter errors,[162] SoundThinking simply never learns of tens of

---

[157] NICK SELBY, DAVID HENDERSON & TARA TAYYABKHAN, SHOTSPOTTER: GUNSHOT LOCATION SYSTEM EFFICACY STUDY 21 (2011), https://perma.cc/4WVG-K77F.

[158] *Id.* at 31.

[159] Garance Burke & Michael Tarm, *Confidential Document Reveals Key Human Role in Gunshot Tech*, AP NEWS (Jan. 20, 2023), https://perma.cc/U7EA-4EUL (documenting a recent claim by SoundThinking that their forensic method has a "97% aggregate accuracy rate for real-time detections across all customers"); City of Chi., *supra* note 143, at 01:26:37, 01:54:43 (during a public hearing on ShotSpotter deployments in Chicago, SoundThinking representatives claimed that their "false positive rate is close to 1%" across the country and also claimed that their overall accuracy rate "over our customer base [is] 97%").

[160] EDGEWORTH ANALYTICS, *supra* note 154, at 2 (emphasis added).

[161] The U.S. Bureau of Justice Statistics reports that about 54% of all violent crime in the United States goes unreported to police. *See* BUREAU OF JUST. STATS., CRIMINAL VICTIMIZATION, 2021, at 1 (U.S. Dep't of Just. 2022), https://perma.cc/2W8M-ZDMH ("About 46% of violent victimizations were reported to police in 2021, higher than in 2020 (40%).").

[162] *See* OIG REPORT, *supra* note 14 (reporting on 37,274 instances (or 89%) of ShotSpotter alerts in Chicago during 2020 and part of 2021 where police responded and encountered no evidence of real gunfire). *See also* ShotSpotter Subpoena Response, State v. Williams,

thousands of instances of potential error. For this unscientific approach of measuring error, these authors simply assume that a lack of feedback means that ShotSpotter must be performing perfectly. For this reason, the accuracy claims in this document provide no serious insights into ShotSpotter performance and do not represent a meaningful substitute for controlled validation testing and error analysis.

The second document's claims that ShotSpotter renders accurate performance and rarely offers erroneous alerts are based on anecdotes from police officers rather than controlled testing. These authors implemented a questionnaire, which included questions such as "What does ShotSpotter do best?"[163] and "In your estimation what percentage of [ShotSpotter] activations are really gunshots and not [other environmental noises]?"[164] Through one-on-one interviews with police personnel, the authors used responses to such questions to report on personal impressions and anecdotes regarding ShotSpotter performance. These survey responses neither provide a meaningful substitute for the empirical validation data described above nor justify the accuracy and performance claims offered by SoundThinking.

IV. INDICATIONS OF SHOTSPOTTER METHOD ERROR IN THE REAL WORLD

Given SoundThinking's inadequate approach to forensic method development and the lack of robust scientific support for their performance claims, failures by their forensic method in the field are inevitable. The strongest indication of classification failure (i.e., the ShotSpotter system's incorrect determination that a sound is or is not a gunshot) comes from operational problems encountered with deployed ShotSpotter systems. Several police departments in the United States have reported on ShotSpotter's false alert problem:

- In 2011, the Broward County (FL) Sheriff's Department discontinued using their $500,000 investment in ShotSpotter after discovering that the system "was wasting too much manpower sending deputies to

No. 20cr0899601 (Ill. Cir. Ct. Cook Cnty. May 21, 2021) (responding to a subpoena demanding the production of "all reclassification notices sent to ShotSpotter by the Chicago Police Department and the Chicago Office of Emergency Management and Communications in 2020" and producing a list of only 27 such notices in 2020) (on file with author).

[163] *See, e.g.*, Telephone Interview by Nick Selby with Brockton (Massachusetts) Police Department (Dispatch), at 12 (Mar. 17, 2011) (on file with author).

[164] *Id.* at 10.

false alarms" generated by firecrackers, car backfires, and other innocent sources.[165]

- In 2012, the Troy (NY) Police Department discontinued ShotSpotter after determining that the microphone system "wasn't reliable," the system generated false alerts from innocent noises on a college campus, and 911 calls did a better job at identifying true gunshot events.[166]
- In 2018, the Chief of Police for Fall River (MA) discontinued ShotSpotter after determining that "ShotSpotter had reported too many false alarms of gunfire while missing actual shot-fired incidents in Fall River."[167] The ShotSpotter system worked "less than 50 percent of the time."[168] The Chief of Police reported that "the city was told that the system was capable 'of doing things it just couldn't do.'"[169]
- In 2016, the Charlotte (NC) Police Department discontinued ShotSpotter after documenting the fact that officers were only "able to find evidence of a gun being fired in one out of 41 reports" at the locations of all ShotSpotter alerts.[170]
- In 2017, the San Antonio (TX) Police Department ended ShotSpotter use after discovering that the system did not work because "[p]olice could find no evidence of a shooting at the scene about 80 percent of the time" and after identifying five shooting victims in ShotSpotter zones that the system failed to detect.[171]

---

[165] Rodriguez, *supra* note 55.

[166] Crowe, *supra* note 63.

[167] Brian Fraga, *After Too Many Shots Missed, ShotSpotter Deal Officially End*, HERALD NEWS, https://perma.cc/3DEG-SPMM (Apr. 20, 2018, 4:59 PM ET).

[168] *Id.*

[169] *Id.*

[170] Cleve R. Wootson Jr., *Charlotte Ends Contract with ShotSpotter Gunshot Detection System*, CHARLOTTE OBSERVER (Feb. 10, 2016, 8:46 PM), https://perma.cc/3JPN-WWPB.

[171] Vianna Davila, *San Antonio Police Cut Pricey Gunshot Detection System*, SAN ANTONIO EXPRESS-NEWS, https://perma.cc/WJ74-2AKL (Aug. 17, 2017, 9:12 AM).

Other comprehensive operational data[172] analyzed by Forbes Magazine[173] and the Chicago Office of Inspector General ("Chicago OIG")[174] is likewise indicative of significant ShotSpotter classification failure. Forbes and the Chicago OIG documented investigative outcomes for tens of thousands of police responses to ShotSpotter alerts, showing that police encountered no evidence of true gun crime events for 89% (37,274 of 41,830) of ShotSpotter alerts.[175] While this data does not translate directly into an 89% false positive rate,[176] the failure of responding police officers to encounter any fired casings,[177] victims, or witnesses at the scenes of most ShotSpotter alerts provides further indications that ShotSpotter has an error problem in need of robust scientific inquiry.

One final indication of an unknown but significant rate of classification error is the level of disagreement on classification decisions between SoundThinking's algorithm and their human examiners. During an eleven-

---

[172] The operational data discussed here is to be distinguished from validation data. With validation data, calculation of meaningful rates of error is possible because ground truth of the testing circumstances in known—the testing is conducted under controlled conditions with known samples. As a matter of definition, ground truth is not known with operational data. For this reason, operational data cannot substitute for validation data and cannot be used to measure true rates of error. Rather, operational data can only be used to provide unscientific insights into possible trends with forensic method performance. *See* Litch & Orrison, *supra* note 151, at 43 (acknowledging that error rate estimations must be based on controlled validation testing but reporting performance trends through the review of operational data, such as the insights that operational data "indicates that during periods such as New Year's Eve and early July the [gunshot detection system under review] is essentially useless to law enforcement" and that false gunshot detection alerts can be caused by exploding transformers, vehicle backfires, thunder, and other sources of impulsive noises); *see also* Mnookin et al., *supra* note 153, at 749 ("Casework may suggest research problems worth exploring. It may lead to hypotheses worth developing. Unusual case findings may be worth discussing at professional meetings or publishing as food for thought. . . . But case findings ought not to be mistaken for structured research or empirical data that goes beyond the anecdotal . . . .").

[173] Matt Drange, *ShotSpotter Alerts Police to Lots of Gunfire, but Produces Few Tangible Results*, FORBES (Nov. 17, 2016, 10:00 AM EST), https://perma.cc/6VQQ-64AC (analyzing operational data from ShotSpotter deployments in Brockton, MA; East Palo Alto, CA; Kansas City, MO; Milwaukee, WI; Omaha, NE; and San Francisco, CA); *see also Short SSTI. ShotSpotter Is Worse Than You Thought*, MOX REPORTS (Nov. 21, 2017), https://perma.cc/889R-32QW.

[174] OIG REPORT, *supra* note 14, at 3.

[175] *Id.*

[176] There are foreseeable explanations for why police may not encounter any evidence—no victims, no witnesses, no fired casings, no bullet holes, no guns—when responding to instances of real gunfire.

[177] Sarah Kollmorgen, *Chicago Criminals' Favorite Gunmakers: A Visual Ranking*, TRACE (Jan. 6, 2016), https://perma.cc/Q2FU-JLC8 (reporting that the three most common gun models used for crimes in Chicago are all semi-automatic handguns, which eject fired casings after each shot).

month period in 2021, SoundThinking's human examiners overrode and changed algorithm classification determinations for noise events in Chicago 94% (648,332 of 691,392) of the time.[178] The only explanations for this level of disagreement between SoundThinking's algorithm and its human examiners are either that SoundThinking's algorithm is highly inaccurate, or that the algorithm is accurate but alerts passed on to police agencies are often rendered inaccurate due to the flawed intervention of SoundThinking's human examiners. Regardless of which is the case, this data shows that the true level of error with ShotSpotter classification decisions—including both steps of algorithm and human analyses—is unknown but concerning.

Separately from these indications of error with ShotSpotter's classification step, known instances of error with ShotSpotter's location estimation process raise additional accuracy questions. For instance, in one high-profile murder case in Rochester, New York, ShotSpotter mislocated the gunfire event by one-and-a-half miles.[179] In other known cases, ShotSpotter was off the mark by 1,400 feet[180] and 130 feet.[181] And in Mr. Williams case in Chicago, ShotSpotter reported two different location estimates that were over one mile apart.[182] Other times, several SoundThinking employees examining the same gunfire event could not agree on where it took place, providing different addresses for the location of the same gunfire event.[183] While the frequency of ShotSpotter location errors in operation is unknown, these reported instances of location-estimation failure point to the need for more robust error testing.

V. OVERSIGHT BY THE SCIENTIFIC COMMUNITY AND THE CRIMINAL JUSTICE SYSTEM

While both the scientific community and the criminal justice system are supposed to play roles in ensuring that rigorous testing and oversight has occurred prior to the use of novel technology in the criminal justice system, neither has played a meaningful role in ShotSpotter oversight yet. With other forensic methods, the scientific community engages in several forms of oversight—creating best practice working groups to generate forensic

---

[178] Letter from Gary Bunyard, *supra* note 144 (based on eleven months of ShotSpotter alert data from Chicago during 2021, ShotSpotter examiners rejected 648,332 (94%) algorithm gunfire classifications and affirmed only 43,060 (6%) algorithm gunfire classifications).

[179] Test. of Paul Greene, *supra* note 47, at 100.

[180] July 6 Testimony in People v. Reed, *supra* note 57, at 227 (450 meters).

[181] *Id.* at 159-60 (40 meters).

[182] Burke et al., *supra* note 4.

[183] July 6 Testimony in People v. Reed, *supra* note 57, at 209-10.

guidelines and standards,[184] developing proficiency and error testing schemes,[185] and creating robust accreditation processes[186]—which are designed to strengthen the scientific footing of the disciplines and increase the chances that evidence offered in the criminal justice system is reliable. Separately, criminal court judges are supposed to act as "gatekeepers,"[187] a role that should require them to comprehensively assess forensic evidence and only admit such evidence in criminal trials when the proponent can conclusively establish that the evidence is borne out of valid forensic methods and the particular results in the case at hand are reliable. In theory, this scientific and legal oversight should represent a significant barrier to the introduction of new and unproven forensic evidence in criminal trials. But in practice, SoundThinking and prosecutors have faced few real barriers when seeking to offer ShotSpotter evidence in criminal prosecutions.

A. The Current Oversight Failure

One important way in which the scientific community provides oversight of forensic methods is to convene subject-matter expert groups to collectively design standards for every step in the forensic analysis process, including evidence handling techniques, analysis methods, reporting limitations, and

---

[184] Administered through the National Institute of Standards and Technology, The Organization of Scientific Area Committees for Forensic Science works to improve forensic science by "facilitating the development and promoting the use of high-quality, technically sound standards." *About OSAC*, NAT'L INST. OF SCI. & TECH., https://perma.cc/5V96-PUAQ.

[185] *See, e.g.*, COLLABORATIVE TESTING SERVICES, INC., https://perma.cc/V6VS-HFMQ (offering annual proficiency testing for ten forensic science disciplines, including DNA, fingerprints, firearms, and others).

[186] *See, e.g.*, AM. NAT'L STANDARDS INST. NAT'L ACCREDITATION BD., ANAB ACCREDITATION FOR FORENSIC SERVICE PROVIDERS 4 (2022), https://perma.cc/9J9S-JPRF (since 1982, ANAB has provided comprehensive accreditation auditing of forensic laboratories designed to assess "a forensic service provider's technical qualifications and competency for conducting specific testing").

[187] Craig Lee Montz, *Judges as Scientific Gatekeepers after Daubert, Joiner, Kumho Tire, and Amended Rule 702: Is Anyone Still Seriously Buying This?*, 33 UNIV. W.L.A. L. REV. 87 (2001) (describing the gatekeeping role assigned to trial court judges in vetting scientific evidence and excluding evidence that is unreliable and not generally accepted in the scientific community).

quality assurance processes.[188] Such vetted standards exist for twenty different forensic science disciplines, from DNA to dog sniffs.[189]

A second important form of scientific oversight involves independent testing of forensic methods and examiners. As described in Section III.C above, independent scientific organizations have conducted and published large-scale error studies that provide important empirics-based method error rate estimations. The scientific community has played important roles in such testing for a long list of forensic disciplines, including DNA,[190] ballistics,[191] shoe print comparison,[192] blood-stain analysis,[193] hair comparison, [194] and bite mark analysis.[195] In addition to these large-scale error studies, the scientific community offers an additional type of forensic testing—examiner proficiency testing.[196] Rather than generating overall rates of error, proficiency testing conducted by independent scientific organizations allows examiners to test their individual competencies on a yearly basis, identify any specific remediation needs, and offer empirical evidence of competency during courtroom testimony.[197] The scientific community has generated proficiency testing for numerous forensic disciplines, including DNA, fingerprints, firearms,

---

[188] The Organization of Scientific Area Committees "was created in 2014 to address the lack of discipline-specific forensic science standards. OSAC fills this gap by drafting proposed standards and sending them to standards developing organizations (SDOs), which further develop and publish them." *About OSAC*, *supra* note 184.

[189] *OSAC Registry,* NAT'L INST. OF SCI. & TECH., https://perma.cc/A65E-6VL6 (listing vetted discipline-wide standards for forensic DNA, arson, firearms and toolmarks, facial recognition, anthropology, dogs and sensors, and other forensic disciplines).

[190] John M. Butler et al., *NIST Interlaboratory Studies Involving DNA Mixtures (MIX05 and MIX13): Variation Observed and Lessons Learned*, 37 FORENSIC SCI. INT'L GENETICS 81 (2018).

[191] DAVID P. BALDWIN, ET AL., A STUDY OF FALSE-POSITIVE AND FALSE-NEGATIVE ERROR RATES IN CARTRIDGE CASE COMPARISONS (2014), https://perma.cc/AH9S-MGY8.

[192] R. Austin Hicklin et al., *Accuracy, Reproducibility, and Repeatability of Forensic Footwear Examiner Decisions*, 339 FORENSIC SCI. INT'L, Oct. 2022.

[193] R. Austin Hicklin et al., *Accuracy and Reproducibility of Forensic Bloodstain Pattern Analysts*, 325 FORENSIC SCI. INT'L, Aug. 2021.

[194] Murrie et al., *supra* note 30, at 2.

[195] D.K. Whittaker et al., *A Comparison of the Ability of Experts and Non-Experts to Differentiate Between Adult and Child Human Bite Marks Using Receiver Operating Characteristic (ROC) Analysis*, 92 FORENSIC SCI. INT'L 11 (1998).

[196] Brendan Max et al., *Assessing Latent Print Proficiency Tests: Lofty Aims, Straightforward Samples, and the Implications of Non-Expert Performance*, 69 J. FORENSIC IDENTIFICATION 281 (2019) (stating that the forensic science and legal communities rely on proficiency testing of forensic examiners to provide overall evidence of method accuracy, as a quality assurance tool in forensic laboratories to identify re-training needs, and as proof of admissibility in court).

[197] *See generally* Brandon L. Garrett & Gregory Mitchell, *The Proficiency of Experts*, 166 U. PENN. L. REV. 901 (2018).

toxicology, forensic handwriting comparison, forensic anthropology, and others.[198] While vetted standards, large-scale error testing, and examiner proficiency testing represent important components of the forensic oversight process, the scientific community has not sought to engage in any of these processes with ShotSpotter, resulting in a significant oversight void.

The criminal justice system likewise has shown little inclination to impose meaningful oversight on ShotSpotter. In theory, criminal court judges should not open courtroom doors to forensic evidence until its admissibility has been litigated at evidentiary hearings in each jurisdiction.[199] In jurisdictions that follow the *Daubert* admissibility scheme, proponents of ShotSpotter evidence have the burden to establish that the method has been properly tested for reliability, has been the subject of peer-reviewed scientific discussion, has generated legitimate error estimations, and has been generally accepted in the scientific community.[200] And of these four *Daubert* factors, the publication of legitimate error rates derived from robust testing is the most concrete and important factor,[201] with some legal commentators suggesting that courts focus the bulk of their *Daubert* analysis on error rate considerations.[202] In the handful of jurisdictions that still apply a general acceptance admissibility standard

---

[198] COLLABORATIVE TESTING SERVICES, INC., *supra* note 185 (stating that CTS is the "first and still the largest forensic proficiency test provider" and that CTS provides "more than 70 tests offered across 10 [forensic] disciplines").

[199] 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 (2006) (explaining that the Supreme Court decision in *Daubert*, followed by the *Joiner* and *Kumho Tire* opinions, stand for "the fundamental principle that trial court judges must act as gatekeepers and carefully screen expert testimony to ensure it reliability.").

[200] Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579 (1993).

[201] John B. Meixner & Shari Seidman Diamond, *The Hidden* Daubert *Factor: How Judges Use Error Rates in Assessing Scientific Evidence*, 2014 WIS. L. REV. 1063 (studying 208 federal *Daubert* decisions and reporting that judges focus on error rate analysis more than other *Daubert* factors).

[202] Munia Jabbar, *Overcoming Daubert's Shortcomings in Criminal Trials: Making the Error Rate the Primary Factor in Daubert's Validity Inquiry*, 85 N.Y.U. L. REV. 2034, 2054, 2057 (2010) ("The error rate factor under the validity inquiry of the *Daubert* standard is the single most important factor that reflects the probative value of expert evidence," and "[i]t is clear that the specific error rate is superior to these other *Daubert* factors as a measure of reliability because peer review and general acceptance remain imperfect proxies for the value of expert evidence."); Jonathan J. Koehler, *Forensics or Fauxrensics? Ascertaining Accuracy in the Forensic Sciences*, 49 ARIZ. ST. L.J. 1369, 1416 (2017) (stating that "the time has surely come for the broader criminal justice system to face the fact that consumers of forensic science evidence (judges, jurors, the public) do not have the information they need to assess the probative value of forensic science opinions and conclusions" and suggesting that scientists and courts focus on error rates in discussions about the validity and admissibility of forensic evidence).

rather than the *Daubert* test, evidence of scientific validity and scientifically-derived error data should also precede admission of forensic evidence.[203]

Despite the fact that legal admissibility schemes should pose real hurdles to the admission of a wide range of forensic and algorithmic evidence,[204] trial court judges in criminal cases have been "utterly ineffective" at vetting forensic evidence.[205] Criminal judges have authorized the admissibility of a laundry list of flawed evidence derived from traditional forensic methods, including hair comparison, bite mark comparison, shoe comparison, bullet lead comparison, arson investigation, and others.[206] More recently with algorithm-based evidence, this same judicial oversight failure has continued. For instance, the proliferation of risk-assessment algorithms commonly relied on by judges during sentencing proceedings has occurred despite limited validation testing and even less testing to establish the rates at which different people operating the same algorithms reach consistent results.[207]

An opinion on the admissibility of evidence from one such risk-assessment algorithm by the Wisconsin Supreme Court provides the clearest example of this judicial failure.[208] In the case, a trial court judge expressly relied on the decision of a risk-assessment algorithm, which labeled a defendant facing sentencing for car theft as a "high risk" for reoffending.[209] The defendant objected to the court's reliance on the algorithm decision, presenting the

---

[203] *See* Geoffrey Stewart Morrison et al., *Consensus on Validation of Forensic Voice Comparison*, 61 SCI. & JUST. 299, 300 (2021) (describing the validation procedures necessary for the production of reliable audio interpretation evidence and asserting that the robust and detailed validation directions therein "describe the consensus as to what is generally accepted within the relevant scientific community.").

[204] Jonathan J. Koehler, *How Trial Judges Should Think About Forensic Science Evidence*, 102 JUDICATURE 28, 36 (2018) ("The problem is not the legal standards pertaining to the admission of forensic evidence" but rather "the failure by courts to take [admissibility standards] seriously.").

[205] COMM. ON IDENTIFYING THE NEEDS OF THE FORENSIC SCIS. CMTY., STRENGTHENING FORENSIC SCIENCE IN THE UNITED STATES: A PATH FORWARD 53 (Nat'l Rsch. Council 2009), https://perma.cc/WJ9P-YFL9 ("In a number of forensic science disciplines, forensic science professionals have yet to establish either the validity of their approach or the accuracy of their conclusions, and the courts have been utterly ineffective in addressing this problem."); Mnookin et al., *supra* note 153, at 734-35 ("Traditional forensic sciences are, at this point, inadequately supported by empirical data that would justify the strong claims analysts frequently make. We believe numerous assertions made both in routine practice and in court are neither backed by sufficient empirical data or research . . . .")

[206] See Saks & Faigman, *supra* note 32, for discussion of vetting failures by criminal court judges.

[207] Brenner et al., *supra* note 11, at 274.

[208] State v. Loomis, 881 N.W.2d 749 (Wis. 2016).

[209] *Id.* at 755.

testimony of an expert who established that, due to the black-box nature of the algorithm, no one involved in the litigation had any understanding of the decision-making process used by the algorithm.[210] In other words, the participants in the case, including the judge, could assess the input data (i.e., information fed into the algorithm) but had no way to assess the output (i.e., the justification for the high-risk classification). The trial judge dismissed those objections and sentenced the defendant to incarceration in prison.[211]

On appeal, the Wisconsin Supreme Court acknowledged that defendants "cannot review and challenge how the [algorithm] calculates risk" because the algorithm outputs "do not explain how the [algorithm] uses information to calculate risk scores."[212] The Court also conceded that the algorithm was not fully validated and that some research indicated that "black defendants were far more likely than white defendants to be incorrectly judged to be at a higher risk of recidivism."[213] And in a frank admission, the Court noted that it could not fulfill its evidence gatekeeping function to assess the validity of the algorithm, stating that "we are not in a position to evaluate or opine on the scientific reliability of this data."[214] Nonetheless, the Court condoned the use of the algorithm in judicial sentencing decision-making, opining that the inability to assess algorithm reliability was sufficiently mitigated by the fact that sentencing judges are instructed to not rely solely on algorithm decisions during sentencing determinations[215] and further taking comfort in the fact that the trial judge and the defendant had an equally limited opportunity to assess the reliability of the algorithm.[216] Based on this reasoning, the highest court in Wisconsin authorized the continued use of a risk-assessment algorithm in critical sentencing determinations without any inquiry into whether risk classifications generated by the algorithm are scientifically defensible.

With ShotSpotter evidence, this same judicial oversight failure has repeated itself. Because of SoundThinking's incomplete approach to algorithm

---

[210] *Id.* at 756-57.

[211] *Id.* at 757.

[212] *Id.* at 761.

[213] *Id.* at 763 (internal quotations omitted).

[214] *Id.* at 762 n.29.

[215] *Id.* at 753 (ruling that the reliance on algorithm output by the sentencing judge was proper because "its use was not determinative in deciding whether Loomis could be supervised safely and effectively in the community.").

[216] *Id.* at 761 ("Additionally, this is not a situation in which portions of [sentencing data] are considered by the circuit court, but not released to the defendant. The circuit court and [the defendant] had access to the same copy of the risk assessment.").

development, lack of peer-reviewed validation data, and complete absence of examiner error data, ShotSpotter proponents lack the necessary bases for carrying their burden under the *Daubert* or general acceptance standards. Nonetheless, ShotSpotter evidence has been admitted in 200 criminal cases in twenty U.S. states.[217]

This widespread acceptance of ShotSpotter evidence by the criminal justice system has occurred, not after robust admissibility litigation, but in the absence of it. Reported judicial opinions in twelve U.S. states and the District of Columbia reference the use of ShotSpotter evidence in criminal trials without any reported ShotSpotter admissibility hearings in those same jurisdictions.[218] Sometimes in these reported decisions, the courts' discussions of SoundThinking's forensic method warranted no more than conclusory footnotes describing SoundThinking's performance claims.[219] Other judges have admitted ShotSpotter evidence in criminal trials without requiring any scientific or legal foundation,[220] mistakenly permitting prosecutors to present ShotSpotter evidence in the form of a written report and in the absence of accompanying expert testimony subject to cross examination.[221] In this way, criminal court judges have approved of ShotSpotter's wide-spread participation in the criminal justice system without the oversight envisioned by the *Daubert* and general acceptance admissibility approaches.

---

[217] *SoundThinking's™ Response to Associated Press Article*, *supra* note 88.

[218] The jurisdictions include Delaware, Georgia, Illinois, Indiana, Louisiana, Massachusetts, Minnesota, New Jersey, Ohio, Pennsylvania, Tennessee, Wisconsin, and the District of Columbia. *See, e.g.*, cases cited *infra* notes 219-20.

[219] Commonwealth v. Mercado, No. 17-P-167, 2018 WL 2089974, at *1 n.3 (Mass. App. Ct. May 7, 2018) (mem.) ("A ShotSpotter is an automated acoustic device used by the Boston police department to detect and locate gunshots."); Commonwealth v. Rafe R., No. 16-P1640, 2018 WL 1023049, at *1 n.3 (Mass. App. Ct. Feb. 23, 2018) (mem.) ("A ShotSpotter is a devise designed to detect gunshots."); Jones v. State, No. 71A04-1507-CR-913, 2016 WL 2983931, at *1 n.1 (Ind. Ct. App. May 24, 2016) (mem.) ("ShotSpotter is an acoustic gunshot detection and location system produced and operated by SST, Inc. that uses microphones in a geographic area to listen for the sound of gunfire. ShotSpotter detects and records the sound of gunfire and uses multilateration . . . to determine the location of the gunfire. It then reports that location to the local law enforcement agencies that are its customers, which here included the South Bend Police Department.")

[220] Commonwealth v. Weeden, 253 A.3d 329, 335 (Pa. Super. Ct. 2021), *appeal granted*, 278 A.3d 305 (Pa. 2022) (holding that prosecutors could seek to admit a ShotSpotter report to prove the existence, location, and timing of a gunshot event through a non-expert police officer witness because "the ShotSpotter report here was automatically generated by the ShotSpotter system and was not an assertion made by a person").

[221] Bullcoming v. New Mexico, 564 U.S. 647 (2011) (holding that it is a violation of the Sixth Amendment's Confrontation Clause to present in a criminal trial forensic evidence in the form of a written report in the absence of the forensic examiner who was directly involved in the forensic analysis).

SoundThinking has played a role in minimizing opportunities for scientific and legal oversight by resisting the basic forms of scientific transparency that are the norm for other forensic method developers. While the U.S. Department of Justice requires forensic labs to make their protocols easily available to any interested party[222] and some forensic laboratories simply post their protocols online,[223] SoundThinking has sought in court to keep their protocol document shielded from disclosure based on claims that disclosure would "jeopardize public safety"[224] and infringe on an "economically valuable" asset of the SoundThinking corporation.[225] Although no forensic labs seek to keep the identity and qualifications of their examiners a secret, SoundThinking routinely argues in court against disclosing this basic information while claiming that such disclosure would have "no bearing on the content or credibility of the [ShotSpotter] evidence."[226] While some forensic algorithm developers post all peer-reviewed validation studies online for easy access by any interested party,[227] SoundThinking seeks to shield their validation from public view by claiming in court that their validation is "quintessentially proprietary" and "is practically a recipe book for the ShotSpotter system."[228] And even though other forensic algorithm developers have provided code access for independent

---

[222] NAT'L COMM'N ON FORENSIC SCI., VIEWS OF THE COMMISSION: ACCREDITATION PROGRAM REQUIREMENT (2016), https://perma.cc/M3N7-TV7P.

[223] *Manuals and Procedures*, VA. DEP'T OF FORENSIC SCIS., https://perma.cc/PM4L-YFDF; *Current Analytical Methods*, IDAHO STATE POLICE FORENSIC SERVS., https://perma.cc/FWG6-TVQD; *Protocols, Procedures, and Validation Summaries*, MICH. STATE POLICE FORENSIC SCI. DIV., https://perma.cc/AXJ4-TYMU; *Technical and Training Manuals,* WASH. STATE PATROL FORENSIC LAB'Y, https://perma.cc/2ENT-SRCE; *Policies*, TENN. BUREAU OF INVESTIGATION, https://perma.cc/GA4C-KPB4; *Policy Manuals and Forms*, CITY OF AUSTIN, https://perma.cc/5LCG-QAB3; *Procedures and Records*, N.C. STATE CRIME LAB'Y, https://perma.cc/G3XM-MJB5; *Department of Forensic Biology*, N.Y.C. OFF. OF THE CHIEF MED. EXAM'R FORENSIC LAB'Y, https://perma.cc/L3TD-Y2C5; *Forensic Science Laboratory Standard Operating Procedures*, D.C. DEP'T OF FORENSIC SERVS., https://perma.cc/7TXB-FL7V.

[224] Third-Party Subpoena Recipient ShotSpotter, Inc.'s Opposition to Defendant's Amended Motion to Modify the ShotSpotter Protective Order at 7, State v. Williams, No. 20cr0889601, (Ill. Cir. Ct. Cook Cnty. May 20, 2021) (on file with author).

[225] *Id.* at 3-6.

[226] Third-Party Subpoena Recipient ShotSpotter, Inc.'s Motion to Quash Subpoenas *Duces Tecum*, *supra* note 77, at 8.

[227] STRmix is an algorithm-based DNA interpretation method which is widely used to generate interpretive findings for admission in criminal litigation. The developers of STRmix make dozens of peer-reviewed validation studies generally available on their website. *See Published Data*, STRMIX, https://perma.cc/7D96-NJB6.

[228] Third-Party Subpoena Recipient ShotSpotter, Inc.'s Motion for a Protective Order and to Quash Subpoenas *Duces Tecum* at 9, State v. Poole, No. 21cr0304701 (Ill. Cir. Ct. Cook Cnty. Nov. 10, 2021).

audits of their algorithms,[229] SoundThinking objects to this form of oversight and has never offered their algorithm for an outside audit.[230] In fact, SoundThinking's refusal to participate in such routine disclosures has resulted in the company being held in contempt of court.[231] Through this approach, which prioritizes secrecy over transparency,[232] SoundThinking operates outside of the scientific norm and hinders the oversight processes.[233]

---

[229] *See, e.g.*, State v. Pickett, 246 A.3d 279, 300 (N.J. Super. Ct. App. Div. 2021) (where the court ordered that the defense be given access to TrueAllele source code for independent evaluation); United States v. Gissantaner, 417 F.Supp.3d 857, 868 (W.D. Mich. 2019) (granting the defense access to STRmix source code); United States v. Johnson, 2016 US Dist. Lexis 194411, at *1 (S.D.N.Y June 7, 2016) (ordering defense access to a proprietary algorithm used by the New York Medical Examiner Office to interpret DNA evidence and commenting that this algorithm was "a relatively new tool that has not been extensively examined or tested in federal court" and noting that the defense could not gain access short of a court order); Order on Defendant's Request to Produce at 30, State v. Conley, No. 48-2012-CT-000017 (Fla. Cir. Ct. 2014), *available at* https://perma.cc/4MT5-4NHQ (granting defense access to the source code for a breath test instrument and holding that "the prosecution cannot proffer evidence and then claim immunity from the obligation to show its evidentiary foundation, especially not on behalf of a private nonparty."); State v. Chun, 943 A.2d 114, 122 (N.J. 2008) (granting defense access to a DUI breath machine despite claims by the manufacturer that such access involved "proprietary information"); United States v. Dioguardi, 428 F.2d 1033, 1038 (2d Cir. 1970) (granting access to a computer algorithm used by an expert witness in forensic accounting and holding that "the defendants were entitled to know what operations the computer had been instructed to perform and to have the precise instruction that had been given").

[230] Third-Party Subpoena Recipient ShotSpotter, Inc.'s Motion to Quash Subpoena *Duces Tecum* in Part at 14-15, People v. Hardy, No. 18cr015233 (Cal. Super. Ct. Alameda Cnty. Feb. 8, 2022).

[231] Matt Chapman & Jim Daley, *ShotSpotter Held in Contempt of Court*, CHI. READER (July 26, 2022), https://perma.cc/ZHN3-SRQJ.

[232] *See* Doucette et al., *Impact of ShotSpotter Technology on Firearms Homicides and Arrests Among Large Metropolitan Counties: A Longitudinal Analysis*, *1999-2016*, 98 J. URB. HEALTH 609, 611 (2021) (reporting that peer-reviewed research on the impact of ShotSpotter systems on crime rates "have been hindered due to the proprietary nature of the data collected by ShotSpotter"). *See also* SHOTSPOTTER, CUSTOMER SUCCESS TRAINING BULLETIN (July 7, 2015) (counseling municipalities to deny access to ShotSpotter information by refusing to comply with routine requests through Public Record Acts and, when forced to comply with such requests, suggesting that municipalities release only "redacted" data to "obscure precise time, location, and rounds fired information"). A screenshot of the bulletin is available at *Short SSTI. ShotSpotter Is Worse Than You Thought*, *supra* note 173 (urging investors to short ShotSpotter shares, in part because of the company's extreme lack of transparency).

[233] *See* KEHL ET AL., *supra* note 11, at 28 ("It is also worth noting the distinction here between algorithms developed by for-profit companies and those created by or in conjunction with non-profits, researchers, and academics. While all of these tools may look like 'black boxes' to outsiders and are susceptible to concerns about opacity, the proprietary tools developed by for-profit companies present unique challenges. Those companies have both an interest in shrouding their products in secrecy in order to remain competitive and more legal tools at their disposal to keep their algorithms away from public scrutiny. Academic researchers and

B. Plan for Robust Oversight

The path to meaningful oversight is clear and involves a shared responsibility between SoundThinking, the scientific community, and criminal court judges. SoundThinking bares primary responsibility for a greater level of openness in the forensic method development process and for initiating required validation testing. The scientific community must also play an active role in the testing process and in other important oversight steps. And the criminal justice system must dust off its admissibility schemes and do the required legal lifting to conduct real vetting of ShotSpotter evidence proffered in criminal cases.

Most importantly, SoundThinking must embrace scientific transparency.[234] The benefits of increased transparency include a better understanding of SoundThinking's black-box method, greater opportunity for outside scientists to audit method performance, and more complete records for ShotSpotter litigation in the criminal justice system.[235] To demonstrate a commitment to scientific transparency, SoundThinking should provide easy access to method protocols, examiner qualifications, and all documentation needed to assess the performance and accuracy of their forensic method. Additionally, SoundThinking should agree to independent audits of their black-box algorithm

governments, by contrast, trend to have more incentives to make the details of their algorithms publicly available and ensure that they are subjected to appropriate scrutiny and oversight.").

[234] Jason M. Chin & Carlos M. Ibaviosa, *Beyond CSI: Calibrating Public Beliefs About the Reliability of Forensic Evidence Through Openness and Transparency*, 62 SCI. & JUST. 272, 281 (2022) ("If forensic science is to maintain its reputation or even improve it, we recommend that forensic scientists critically examine their research and practices and consider aligning them with the move towards openness and transparency occurring elsewhere within science. In this path, credibility is earned because the research and work underlying forensic science claims are transparently reported such that they can be tested by other scientists. [] Openness and transparency also help ensure that forensic scientific evidence comports with the rules of evidence and the principles behind them."); NAT'L ACAD. SCI., ENG. & MED., OPEN SCIENCE BY DESIGN: REALIZING A VISION FOR 21ST CENTURY RESEARCH 4 (2009) ("Research conducted openly and transparently leads to better science. Claims are more likely to be credible—or found wanting—when they can be reviewed, critiqued, extended, and reproduced by others."); Mnookin et al., *supra* note 153, at 743 ("A research culture maximizes transparency, both in the production of knowledge and in internal practices and in internal practices and procedures. Researchers should be encouraged to make data sets available to other researchers, both to share the particular basis for their own claims and to encourage further research.").

[235] KEHL ET AL., *supra* note 11, at 32 ("While transparency alone will not necessarily reduce the likelihood of [algorithmic] bias, it remains valuable for a number of reasons. First and foremost, greater transparency can help facilitate audits by outside researchers. It can also help increase the general understanding of these systems, how they work, and the tradeoffs involved in implementing them.").

and the datasets used to train it.[236] In providing access to their underlying computer code and data, SoundThinking can effectively maintain any proprietary interests they have in their technology through straightforward legal processes designed to protect intellectual property rights.[237]

In addition to increased transparency, both SoundThinking and the scientific community must engage in robust peer-reviewed[238] validation and error analysis of SoundThinking's forensic method.[239] SoundThinking can start this process by conducting and publishing the type of validation testing described above, including robust internal validation testing for each location where ShotSpotter is deployed. To ensure objectivity and transparency of their validation testing, SoundThinking should embrace the best practice of pre-registration, which involves public disclosure of validation test planning documentation prior to the start of validation testing.[240] Separately, the

---

[236] *Cf.* cases cited *supra* note 229 (examples of courts ordering proprietary algorithms and source code to be provided to criminal defendants for independent examination).

[237] Rebecca Wexler, *Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System*, 70 STAN. L. REV. 1343, 1409-10 (2018) ("For trade secret evidence that satisfies the criminal discovery or subpoena requirements, courts can mitigate any risk from disclosure by using protective orders, sealing orders, and limited courtroom closures. . . . In civil discovery, judges routinely order trade secrets disclosed to opposing parties under protective orders.").

[238] Mnookin et al., *supra* note 153, at 744 ("Research projects should be designed according to the norms of relevant academic fields. They should not be designed defensively, to produce, or to increase the chances of producing, a particular outcome. Publication and peer review should occur as a matter of course, and a commitment to publication should not depend on the results.").

[239] *Id.* at 51 (a diverse group of commentators, including practicing forensic examiners as well as critics, agreeing that "many forms of forensic science today stand on an insufficiently developed empirical research foundation" and recommending that the forensic science community needs "to increase their commitment to empirical evidence as the basis for their claims").

[240] Chin & Ibaviosa, *supra* note 234, at 279-80 ("[F]orensic science researchers should consider preregistering their research on one of many public registries. Preregistration involves placing a timestamped version of a research protocol online . . . prior to the data being collected. Unlike traditional methods of conducting research, this combats the file drawer problem . . . because even if the study is never formally published, other researchers can find it and include it in meta-analyses and systematic reviews . . . . Preregistration also improves transparency in that others can compare the final published protocol to what was initially planned to see what changes were made after the data was collected . . .—and to see if any deviations from the initial preregistration were sufficiently justifiable."); Brian A. Nosek et al., *The Preregistration Revolution*, 115 PNAS 2600, 2605 (2018) ("Sometimes researchers use existing observations of nature to generate ideas about how the world works. This is called postdiction. Other times, researchers have an idea about how the world works and make new observations to test whether that idea is a reasonable explanation. This is called prediction. To make confident inferences, it is important to know which is which. Preregistration solves this challenge by requiring researchers to state how they will analyze

scientific community should devise the types of examiner competency and large-scale error testing that have been implemented in most other forensic disciplines.[241] Importantly, this testing must be implemented blindly,[242] meaning that SoundThinking examiners should be "unaware they are being tested."[243] Blind error testing is essential because examiners tend to change their behavior when they know they are being tested, leading to flawed estimations of examiner error rates.[244] Blind testing of SoundThinking examiners is especially important in light of the fact that SoundThinking examiners' behavior during casework may be biased by the built-in incentives in ShotSpotter contracts with local municipalities.[245] Other forensic labs have

---

the data before they observe it, allowing them to confront a prediction with the possibility of being wrong. Preregistration improves the interpretability and credibility of research findings.").

[241] Error testing should assess at least four critical metrics regarding ShotSpotter's classification performance: the false negative rate, the false positive rate, the rate of reproducibility, and the rate of repeatability. *See* Kori Khan & Alicia L. Carriquiry, *Shining a Light on Forensic Black Box Studies*, ARXIV (Sept. 28, 2022, 4:42 PM UTC), https://perma.cc/4JRU-ZBK8 (explaining that each of these performance metrics are commonly included in error testing in other forensic disciplines).

[242] Robin Mejia et al., *Implementing Blind Proficiency Testing in Forensic Laboratories: Motivation, Obstacles, and Recommendations*, 2 FORENSIC SCI. INT'L: SYNERGY 293, 293 (2020) ("Blind proficiency tests involve samples that are submitted through the normal analysis pipeline as if they were real cases, requests, or tenders. In blind tests, the examiners conduct the analysis under the assumption they are working on real samples. Only after the work is completed do they learn that a case was a proficiency test.").

[243] Brett O. Gardner et al., *Latent Print Quality in Blind Proficiency Testing: Using Quality Metrics to Examine Laboratory Performance*, 324 FORENSIC SCI. INT'L, July 2021, at 1 (reporting that non-blind proficiency tests "do not generalize to real-world casework because analysts' test-taking behavior is not representative of routine casework").

[244] Mejia et al., *supra* note 242, at 294 ("Studies from other testing industries have shown that both behavior and results can differ when examiners are given declared v. blind proficiency tests. Two studies in drug testing labs in the 1970s compared blind and declared proficiency tests at 24 and 10 labs, respectively, and found that false negatives were higher in the blind tests compared to when laboratories knew they were being tested. [] A 2001 study comparing blind and declared proficiency tests in blood lead testing programs at two large state laboratories found error rates were highest in the blind tests and suggested that laboratories were making special efforts when analyzing known proficiency test samples. Today, the Mandatory Guidelines for Federal Workplace Drug Testing Programs require participating laboratories to conduct blind testing.").

[245] The contract between SoundThinking and the City of Chicago only subjects SoundThinking to a monetary penalty when examiners fail to report real gunfire events, not when examiners falsely report innocent noises as gunfire. Contract between ShotSpotter, Inc. d/b/a SST, Inc. and City of Chicago (Aug. 20, 2018), *available at* https://perma.cc/WY4K-2MU9.

implemented such blind testing[246] and it represents the gold standard in examiner testing.[247]

Even though attaining accreditation is not a panacea for poor forensic method development and implementation, SoundThinking should nonetheless be required to seek forensic accreditation. Forensic accreditation involves a comprehensive audit by outsiders who review lab practices for compliance with fundamental attributes of good science, including (1) validated methods, (2) written protocols, (3) defined quality assurance processes, and (4) documentation of error remediation.[248] Through the accreditation audit process, forensic laboratories can demonstrate their "compliance to industry standards" as well as their "capacity to generate and interpret results."[249] Because there is "little doubt that universal accreditation of forensic science service providers would have a salutary impact both on the validity of forensic testing and the level of public trust in the forensic evidence brought to bear in the courtroom,"[250] numerous scientific and legal organizations—including the National Commission on Forensic Science,[251] the American Bar Association,[252] and other groups[253]—deem accreditation a critical step for all forensic science

---

[246] *E.g.*, The Houston Forensic Science Center (HFSC). *See* Mejia et al., *supra* note 242, at 295 (discussing how HFSC is a leader in blind testing, digital forensics, latent prints, toxicology, and other forensic disciplines); Maddisen Neuman et al., *Blind Testing in Firearms: Preliminary Results from a Blind Quality Control Program*, 67 J. FORENSIC SCI. 964, 972 (2022) (reporting the results of a multi-blind testing program for forensic firearms examiners and concluding that blind testing procedures "allow for a more accurate and effective measure of how examiners and processes and procedures are operating").

[247] Garrett & Mitchell, *supra* note 197, at 959 ("Only by demanding data from realistic blind proficiency testing will courts ensure that parties and their experts come forward with the data needed to ensure that an expert truly is an expert. In mandating this information, judges will greatly simplify the question of expert admissibility, avoiding the more complex methodological inquires called for by Rule 702 and *Daubert*.").

[248] *See* LUDWIG HUBER, UNDERSTANDING AND IMPLEMENTING ISO/IEC 17025, at 9 (2009), https://perma.cc/VYY7-68L7.

[249] NAT'L COMM'N ON FORENSIC SCI., VIEWS OF THE COMMISSION: CRITICAL STEPS TO ACCREDITATION 2 (2016), https://perma.cc/2286-BRTV.

[250] SUBCOMM. ON FORENSIC SCI., STRENGTHENING THE FORENSIC SCIENCES 4 (Nat'l Sci. & Tech. Council 2014), https://perma.cc/K6SV-WM54.

[251] NAT'L COMM'N ON FORENSIC SCI., UNIVERSAL ACCREDITATION 2, https://perma.cc/7JH3-WR8A (stating that "[t]o improve the quality of forensic science, all entities performing forensic science testing, even on a part-time basis, must be included in universal accreditation").

[252] A.B.A., ABA STANDARDS FOR CRIMINAL JUSTICE: DNA EVIDENCE 5 (3d ed. 2007), *available at* https://perma.cc/2MUM-R2S7 (recommending that all "laborator[ies] testing DNA evidence should: be accredited every two years under rigorous accreditation standards by a nonprofit professional association actively involved in forensic science and nationally recognized").

[253] *See, e.g.*, EXPERT WORKING GRP. ON HUM. FACTORS IN LATENT PRINT ANALYSIS, LATENT PRINT

laboratories. In demanding that SoundThinking attain forensic accreditation, the scientific and legal communities can nudge SoundThinking toward the implementation of the good scientific practices that they have not implemented on their own.

As the ultimate gatekeepers of forensic evidence in the criminal justice system, criminal court judges should step up and provide robust vetting of ShotSpotter evidence. According to one group of influential forensic science commentators, continued judicial inaction will only encourage the use of substandard forensic evidence in the criminal justice system:

> If courts are not going to insist upon better evidence of validity, if they are instead going to continue to permit forensic scientists to reach extremely strong conclusions about their own abilities to make identifications, and if legal challenges remain both relatively rare and generally unsuccessful, then why should the forensic science community consider changing its practices?[254]

If, however, criminal court judges decide to take their role of forensic evidence gatekeeper seriously, the *Daubert* admissibility scheme provides them with the tool necessary to engage in the robust oversight needed with ShotSpotter evidence. To inject this tool with some legal efficacy, judges should require that prosecutors seeking to use ShotSpotter evidence present sufficient scientific evidence that: (1) SoundThinking's method has been fully validated, (2) their method has been subject to peer-reviewed scrutiny, (3) SoundThinking has generated science-derived error rates, (4) SoundThinking has implemented science-derived standards for the proper use of their method, and (5) multiple independent large-scale studies have assessed the performance of SoundThinking's human examiners.[255] With regard to full validation, criminal

---

EXAMINATION AND HUMAN FACTORS: IMPROVING THE PRACTICE THROUGH A SYSTEMS APPROACH (2012), https://perma.cc/TYA8-YLUY; COMM. ON IDENTIFYING THE NEEDS OF THE FORENSIC SCIS. CMTY., *supra* note 205, at 76; Mnookin et al., *supra* note 153, at 733 (recommending "[m]andatory accreditation of all forensic science laboratories that process evidence for court").

[254] Mnookin et al., *supra* note 153, at 758-59.

[255] Even in jurisdictions that apply the Frye general acceptance standard to admissibility questions involving forensic evidence, trial court judges must nonetheless require robust evidence of method validation and accuracy, including science-derived false positive and false negative error rates. *See* Bader v. Johnson & Johnson, 303 Cal. Rptr. 3d. 162, 201 (Cal. Ct. App. 2022) (Streeter, J., concurring) (stating that Frye admissibility determinations should include analysis of whether a scientific theory "is testable by 'empirical demonstration of accuracy'" and suggesting that "among the most important criteria for testable empirical

court judges should demand that ShotSpotter proponents produce evidence of robust compliance with all three steps in the validation process: developmental, internal, and V&V. Regarding error rates, judges should exclude ShotSpotter evidence until such time that proponents can produce multiple peer-reviewed blind studies on controlled samples that empirically quantify the rates of error, including the false positive error rate, for SoundThinking's human examiners as well as their repeatability and reproducibility rates.[256] On rare occasions, courts have applied such a robust admissibility analysis to novel prosecution-proffered forensic evidence.[257] This approach needs to be the norm rather than the exception.[258]

VI. THE RACIAL IMPLICATIONS OF CONTINUED SHOTSPOTTER DEPLOYMENTS

The importance of robust scientific and legal oversight of ShotSpotter tech is heightened by the role SoundThinking plays in perpetuating racial inequalities in the criminal justice system. Algorithm-based discrimination occurs when "automated systems contribute to unjustified different treatment or impacts" that disfavor groups of people due to their race, ethnicity, sex, or other identifiers.[259] Algorithms have been shown to perpetuate inequality in

---

accuracy is whether 'error rates' have been taken into account"); *see also* People v. McKown, 924 N.E.2d 941 (Ill. 2010) (conducting a Frye analysis to assess whether the horizontal gaze nystagmus testing is a generally accepted indicator of alcohol impairment and weighing peer-reviewed scientific literature and expert testimony regarding the validated uses of the method and its known rates of error).

[256] *See* PRESIDENT'S COUNCIL OF ADVISORS ON SCI. & TECH., *supra* note 34, at 47-54 for a discussion of the meaning of repeatability and reproducibility in the validation context.

[257] Williamson v. Reynolds, 904 F. Supp. 1529 (E.D. Okla. 1995) involves the rare exclusion of prosecution proffered forensic evidence. In excluding hair comparison evidence pursuant to *Daubert*, the court rigorously applied the *Daubert* factors to assess admissibility. Applying the first *Daubert* factor of method validity, the court noted that the practice of forensic hair comparison lacked critical method standards, including "accurate definitions of hair features in microscopic hair examination." *Id.* at 1554 n.11. Regarding the second *Daubert* factor of peer-reviewed publication, the court noted the "apparent scarcity of scientific studies regarding the reliability of hair comparison testing." *Id.* at 1556. When discussing the error rate factor, the court noted the paucity of such studies and cited to an existing study which reported that "error rates on hair analysis were as high as 67%." *Id.* In analyzing the *Daubert* factor of general acceptance, the court noted that "any general acceptance seems to be among hair experts who are generally technicians testifying for the prosecution, not scientists who can objectively evaluate such evidence." *Id.* at 1558.

[258] Mnookin et al., *supra* note 153, at 761 ("If, for example, courts insisted on better error rate information as a precondition for admissibility, the incentives for its production would dramatically increase.").

[259] WHITE HOUSE OFF. OF SCI. AND TECH. POL'Y, BLUEPRINT FOR AN AI BILL OF RIGHTS: MAKING AUTOMATED SYSTEMS WORK FOR THE AMERICAN PEOPLE 5 (2022), https://perma.cc/KZX5-Y5WK.

numerous contexts, including medical care, banking, and employment.[260] In the criminal justice system, entrenched inequalities are now exacerbated by the wide-spread adoption of ShotSpotter technology. SoundThinking's role in exacerbating existing inequalities stems from the fact that ShotSpotter alerts result in thousands of people in communities of color undergoing unjustified police suspicion and investigation in a manner that does not occur in majority White communities.

Extensive research has "documented substantial racial and ethnic disparities at each stage of the criminal justice process."[261] One unabating factor driving this racial inequality is the disproportionate rate at which police officers stop and search people of color.[262] Black Americans are both more likely to be subjected to police-initiated stops than White Americans[263] and more likely to suffer violence at the hands of police during these encounters.[264] Even when initial police encounters do not lead to formal arrests, Black Americans are nonetheless more likely to be seen by police as "guilty until proven otherwise"[265] and subjected to more disrespectful police behavior during these encounters.[266] The deployment of ShotSpotter systems in the United States has perpetuated racial inequalities in the criminal justice system by generating an

---

[260] *Id.* at 3 ("In America and around the world, systems supposed to help with patient care have been proven unsafe, ineffective, or biased. Algorithms used in hiring and credit decisions have been found to reflect and reproduce existing unwanted inequalities or embed new harmful bias and discrimination. Unchecked social media data collection has been used to threaten people's opportunities, undermine their privacy, or pervasively track their activity—often without their knowledge or consent. These outcomes are deeply harmful—but they are not inevitable.").

[261] NAT'L ACADS. OF SCI., ENG'G & MED., REDUCING RACIAL INEQUALITY IN CRIME AND JUSTICE: SCIENCE, PRACTICE AND POLICY 1 (2023), *available at* https://perma.cc/C3U3-HB5J.

[262] *Id.* at 3, 6 ("Police officers stop and search Black individuals at rates that are higher than for other racial and ethnic groups. . . . [T]he early stages of the system—including police stops, jail confinement, misdemeanor courts, and fines and fees—generate vast numbers of contacts (relative to White communities) between police and courts on the one hand and Black, Latino, and Native American communities on the other.").

[263] *Id.* at 66 (reporting that most such stops do not result in the finding of criminal activity resulting in arrests).

[264] *Id.* ("[A]lthough police rarely use force during stops, they are more like to use force when they stop African Americans, even when the stop does not begin because police believe that a crime is in progress.").

[265] *Id.* at 160.

[266] *Id.* at 81 (describing a study which assessed police conduct as recorded on body-worn cameras found that "officers speak with consistently less respect toward Black versus White community members" and describing a second study which found that "racial disparities in intonation [of police officer voices during street encounters] undermine trust in institutions such as police departments").

overwhelming number of ShotSpotter-initiated police investigations in communities of color.

The use of ShotSpotter in Chicago provides one example of its disparate racial impact. ShotSpotter is deployed along racial lines in Chicago: systems are deployed in every police district where people of color comprise at least 65% of residents and are not deployed in any police district where the majority of residents are White.[267] Once deployed, ShotSpotter systems initiate tens of thousands of police investigations annually in these communities of color, including 50,176 investigations during a seventeen-month period in 2020 and 2021.[268]

While the vast majority (89%) of these ShotSpotter-initiated police actions resulted in police encountering no evidence of a real gun crime and even fewer (less than 1%) resulted in the recovery of a firearm, police responses to these alerts nonetheless resulted in over one thousand pat-downs and searches of people in these communities.[269] And even when police did not discover evidence of real gun crimes, they often (948 times in seventeen months) used ShotSpotter alerts as the justification for initiating stops of people, which resulted in arrests for a host of allegations unrelated to gunfire, including Reckless Conduct, Interference with a Public Officer, Obstruction of Justice, various narcotics possession offenses, and other criminal charges.[270] In this way, the increased number of police encounters and searches generated by ShotSpotter in communities of color perpetuates existing racial disparities in policing, resulting in an added layer of police encounters which does not occur in majority White neighborhoods.

In light of these concerns about ShotSpotter's disparate racial impact, scientists should conduct a formal bias impact study[271] to scrutinize the racial implications of ShotSpotter deployments. Such a study should report on which communities are covered by ShotSpotter deployments, the extent to which deployments may contribute to existing historical racial inequities in arrest and conviction rates among people of color, and mitigation strategies to address

---

[267] State v. Williams Amicus Brief, *supra* note 38, at 14.

[268] OIG REPORT, *supra* note 14, at 3.

[269] OIG REPORT, *supra* note 14, at 3, 16.

[270] OIG REPORT, *supra* note 14, at 24.

[271] Lee, Resnick & Barton, *supra* note 22 (defining a bias impact statement in the context of algorithm development as a formal "self-regulatory practice" used to assess any racial impacts resulting from the purpose, production, and deployment of algorithms in society).

racial impacts.[272] Conducting bias impact studies is a best-practice step in the algorithm development process,[273] and by demanding that ShotSpotter undergo such an analysis alongside comprehensive validation and error testing, the criminal justice system can take a small step toward the aspiration of justice for all.

CONCLUSION

The deployment of ShotSpotter tech across the United States and the routine use of ShotSpotter evidence in criminal cases have occurred in the absence of robust empirical evidence of scientific performance and societal impact. For a criminal justice system that has a history of failing to provide meaningful gatekeeping of forensic evidence, this failure is predictable but avoidable.[274]

---

[272] CHRISTOPHE ABRASSART ET AL., MONTRÉAL DECLARATION FOR A RESPONSIBLE DEVELOPMENT OF ARTIFICIAL INTELLIGENCE 6, 13 (2018), https://perma.cc/B674-XQP3 (offering ten specific guidelines for ethical algorithmic development and usage "born from an inclusive deliberation process" and reporting the principle that algorithms "must be designed and trained so as not to create, reinforce, or reproduce discrimination"); Lee, Resnick & Barton, *supra* note 22 ("In the decision to create and bring algorithms to market, the ethics of likely outcomes must be considered—especially in areas where governments, civil society, or policymakers see potential for harm, and where there is risk of perpetuating existing biases or making protected groups more vulnerable to existing societal inequalities. That is why it's important for algorithm operators and developers to always be asking themselves: *Will we leave some groups of people worse off as a result of the algorithm's design or its unintended consequences?*" (emphasis in original)).

[273] Lee, Resnick & Barton, *supra* note 22 ("As a self-regulatory practice, the bias impact statement can help probe and avert any potential biases that are baked into or are resultant from the algorithmic decision. As best practice, operators of algorithms should brainstorm a core set of initial assumptions about the algorithm's purpose prior to its development and execution. We propose that operators apply the bias impact statement to assess the algorithm's purpose, process and production, where appropriate."); *see also* Ben Shneiderman, *The Dangers of Faulty, Biased or Malicious Algorithms Requires Independent Oversight*, PNAS (Nov. 29, 2016), https://perma.cc/7DTP-WTA3 ("When major new or revised algorithm systems are being developed, an independent oversight review could require implementers to submit an algorithm's impact statement. This document would be similar to the environmental impact statements that are now required for major construction programs."). One major organization implementing this practice is the IEEE, which is the largest professional organization for computer engineers in the world. Over several years, the IEEE convened a panel of computing scientists to consider guidelines for the ethical development of computer algorithms. Among their recommendations, the IEEE recommends that the "[e]valuation of [algorithms] must carefully assess potential biases in the system's performance that disadvantage specific social groups. This evaluation process should integrate members of potentially disadvantaged groups to diagnose and correct such biases." INST. OF ELEC. & ELECS. ENG'RS, *supra* note 29, at 52.

[274] For example, the judiciary has demonstrated an ability to apply robust admissibility oversight to scientific evidence in civil cases. *See* sources cited *supra* note 32.

This picture will not change on its own.[275] It will only improve with sufficient engagement and oversight from the scientific and legal communities to determine whether deployment of ShotSpotter systems in communities of color serves defined purposes while not imposing harm on the people who live in those communities.[276] By addressing head-on SoundThinking's substandard approach to forensic method development and requiring a more robust scientific foundation, the scientific and legal communities can recognize that faulty algorithms in the criminal justice system are public problems and not just legal admissibility questions impacting individual defendants.[277] Until this realization leads to real oversight and vetting, people of color like Michael Williams will continue to suffer the impacts[278] of routine ShotSpotter-initiated police encounters.

---

[275] *See* INST. OF ELEC. & ELECS. ENG'RS, *supra* note 29, at 60 ("Corporations, whether for-profit or not-for-profit, are eager to develop, deploy, and monetize [algorithms], but there are insufficient structures in place for creating and supporting ethical systems and practices around [algorithmic] funding, development, or use.").

[276] *Cf.* INST. OF ELEC. & ELECS. ENG'RS, *supra* note 29, at 3 ("As the use and impact of [algorithms] become pervasive, we need to establish societal and policy guidelines in order for such systems to remain human-centric, serving humanity's values and ethical principles. These systems have to behave in a way that is beneficial to people beyond reaching functional goals and addressing technical problems. This will allow for an elevated level of trust between people and technology that is needed for its fruitful, pervasive use in our daily lives.").

[277] Mike Ananny, *Seeing Like an Algorithmic Error: What Are Algorithmic Mistakes, Why Do They Matter, and How Might They Be Public Problems*, 24 YALE J.L. & TECH. 1, 5 (2022) ("When algorithmic errors are public problems, they are not idiosyncratic quirks for software companies to debug privately and on their own timeline. They are instead powerful provocations showing—exactly—how a system has failed, why it has failed, what its successful operation would look like, who benefits from its failures, and how reformers can fix the mistake, remedy the harms, and prevent future errors.").

[278] Complaint at 62-67, Williams v. City of Chicago, No. 1:22-cv-03773 (N.D. Ill. July 21, 2022), ECF No. 1, *available at* https://perma.cc/FXJ9-LUAV (civil class action complaint alleging the ways in which Michael Williams' ShotSpotter-initiated wrongful incarceration impacted him, including worsening health crises, an inability to access medications and medical care, financial instability, and thoughts of suicide).