11 July 2022

National Institute of Standards and Technology (NIST)
U.S. Department of Commerce
100 Bureau Drive
Gaithersburg, MD 20899
Via Email: scientificfoundationreviews@nist.gov

Re: *RFC Response: Digital Investigation Techniques: A NIST Scientific Foundation Review (NISTIR 8354-DRAFT)*

IEEE is pleased to submit these comments on the above-captioned, Request for Comment on NIST's *Digital Investigative Techniques: A NIST Scientific Foundation Review* (8354-DRAFT, "the Review"). These comments are the combined effort of IEEE-USA, the IEEE Standards Association (IEEE SA), and the IEEE Computer Society.

IEEE-USA is the American component of the IEEE, representing approximately 150,000 engineers, scientists, and allied professionals in United States, many of whom are actively conducting research and development into digital forensics, privacy, cybersecurity, artificial intelligence, software engineering, and advanced computing, as well as other foundational and emerging technologies.[1]

The IEEE SA is a globally recognized standards-setting body within IEEE.[2] The IEEE standards development process is rooted in consensus, due process, openness, right to appeal and balance.[3] It adheres to and supports the principles and requirements of the World Trade Organization's (WTO) Decision on Principles for the Development of International Standards, Guides and Recommendations. In particular, the IEEE operates in active agreement with the WTO principle that standards should not create unnecessary obstacles to trade, and whenever appropriate, should specify requirements in terms of performance rather than design or descriptive characteristics.[4]

The IEEE Computer Society (Computer Society) is the premier source for information, inspiration, and collaboration in computer science and engineering.[5] The Computer Society has over 373,000 members from 168 countries; hosts 215 technical conferences annually; maintains 230 active global technical standards; and, publishes 47 peer-reviewed journals and magazines including those that the Review[6] lists as leading publications on digital forensics and computer security (e.g., IEEE Transactions on Cloud Computing, IEEE Transactions on Information Forensics and Security, and IEEE Transactions on Software Security).

Our society sits at the intersection between enabling new levels of technological efficiency and enabling technology to become a force for good that goes beyond efficiency. We have a critical opportunity to use technology to make society more equitable, inclusive, and just; make government operations more transparent and

[1] https://ieeeusa.org/
[2] https://standards.ieee.org/
[3] "Developing standards: Who Oversees The Process?" IEEE Standards Association. https://standards.ieee.org/develop/develop-standards/govern/ (defining Due Process as "having highly visible procedures for standards creation and following them;" Openness as "ensuring that all interested parties can participate and are not restricted to a particular type or category," Consensus as requiring a supermajority of a group to approve a draft standard (75% of the ballots must be returned, with 75% of them voting yes); Balance as "ensuring that voting groups include all interested parties.
[4] IEEE, "IEEE Adherence to the World Trade Organization Principles for International Standardization," August 19 2020. Available at: https://globalpolicy.ieee.org/wp-content/uploads/2020/08/IEEE20013.pdf
[5] https://www.computer.org/about
[6] Review, § 2.8.8

IEEE-USA | 2001 L Street, N.W., Suite 700, Washington, D.C. 20036-4928 USA
Office: +1 202 785 0017 | Fax: +1 202 785 0835 | E-mail: ieeeusa@ieee.org | Web: http://www.ieeeusa.org

accountable; and encourage public participation and increase the public's trust in government. When used according to these objectives, technology can help reaffirm and protect our democratic values.

If we miss the opportunity to use these technologies to ensure protection of human values and trustworthiness, we risk reinforcing disparities in access to goods and services, discouraging public participation in civic life, and eroding the public's trust in government. Put another way: responsible development and use of technology to further safeguard human values and ensure trustworthiness is an approach that leads to a sustainable ecosystem of innovation. It is this type of approach that our society will trust and accept.

To achieve the stated objectives we draw upon our collective expertise and provide the following recommendations to address the needs specified in the draft of the scientific foundation review.[7] To summarize,

1. Digital forensics and their results should be deemed reliable based on objective information gathered through independent verification and validation as informed by standards.
2. Digital forensics and their results should be deemed reliable based on objective information gathered through independent verification and validation standards.
3. Digital forensics should be tested against and governed by standards adhering to principles of due process, openness, consensus, balance, and right of appeal.
4. To ensure digital forensics' effectiveness, competence, awareness, accountability, and transparency in operation, there must be standards and certifications for digital forensics and their operators, and recurring benchmarking exercises and independent studies.
5. Determining the reliability and trustworthiness of forensic technologies like digital forensics requires evaluating them in their operational environments, their use in legal proceedings and how fit the technology is for those uses.
6. Stakeholders of digital forensics must have appropriate access to the software necessary to assess the degree of reliability and validity of information, and whether that information is fit-for-purpose.
7. Trustworthiness is determined by more than reliability, and therefore, requirements to determine trustworthiness must include assessments of the processes and procedures where these systems are deployed.
8. To ensure digital forensics are reliable and trustworthy, governments should provide sufficient funding for testing, evaluation, certification, and investigation of digital forensics.
9. NIST's own Text Retrieval Conference (TREC) program should be used as an example to inform NIST policy in otherareas such as digital forensics.

We expand on each of these 9 recommendations in detail below. In particular, we make specific and substantial suggestions for changes to KEY TAKEAWAY #2.5, #4.2, #4.4, #4.5, #4.6 and #4.7 in the document.

We thank NIST for considering these comments in the agency's revisions to the Request for Comment on NIST's Digital Investigation Techniques: A NIST Scientific Foundation Review (NISTIR 8354-DRAFT). We would welcome any further discussions with the agency on these matters. If you have questions, please do not hesitate to contact Erica Wissolik at (202) 530-8347 or e.wissolik@ieee.org.

[7] For a similar statement, see IEEE-USA "Letter to the National Institute of Standards and Technology (NIST), responding to request for comments on NIST Internal Report 8351-DRAFT *DNA Mixture Interpretation: A NIST Scientific Foundation Review*," November 18, 2021. Available: https://www.nist.gov/document/nist-ai-rfi-ieee001pdf [Accessed July 7, 2022].

RECOMMENDATIONS

1. **Digital forensics and their results should be deemed reliable based on objective information gathered through independent verification and validation as informed by IEEE 1012(™), Standard for System, Software, and Hardware Verification and Validation. Without that objective information, neither these systems, nor their results can be considered reliable or trustworthy.**

The use of digital forensics in criminal court can result in catastrophic failures through false imprisonment and the deprivation of people's rights. Scientists and engineers have long demanded that safety-critical software and hardware be the right systems built the right way. Therefore, digital forensics should be independently verified and validated (IV&V) prior to deployment, or prior to informing decisions in the legal system, law enforcement, governance, and related compliance. Specifically, digital forensics ought to be independently verified and validated in accordance with technical standards such as IEEE 1012 Standard for System, Software, and Hardware Verification and Validation,[8] and be subject to recurring post-deployment audit, including with respect to their operators. We encourage NIST to uphold these same requirements.

IEEE 1012 provides a universally applicable and broadly accepted process for helping to ensure that a product is correctly built for its intended use. For example, it is used to verify and validate Department of Defense nuclear weapons systems and NASA manned space systems and critical space exploration probes, among many others.[9]

IV&V are interrelated and complementary processes that build quality into any system. Verification is focused on a product, providing objective evidence for whether the product conforms to requirements, standards, and practices. Validation is focused on customers and stakeholders, providing evidence for whether a product is accurate and effective, solves the right problem, and satisfies the intended use and user needs in the operational environment. In short, verification ensures that a product is correctly built, while validation ensures that the right product is built.

In the context of digital forensics, IV&V answers the following types of questions: Is the analysis used by the digital forensics software the best available, coded as designed, and appropriate for the problem? Does digital forensics systematically favor including or excluding certain types of information? How likely are false negatives and false positives? Would outside experts agree with the software's results at each stage of analysis?

To help appropriately perform IV&V, IEEE 1012 requires that each software and hardware component be assigned an integrity level that increases depending on the likelihood and consequences of a failure: negligible, marginal, critical (causing "major and permanent injury, partial loss of mission, major system damage, or major financial or social loss," referred to as integrity level 3), and catastrophic (causing "loss of human life, complete mission failure, loss of system security and safety, or extensive financial or social loss," referred to as integrity level 4).[10] As the integrity level increases, so too does the intensity and rigor of the required IV&V tasks.

Digital forensics analysis tools should undergo IV&V according to its integrity level as defined by IEEE 1012. Because a thorough and public conversation is yet to take place, there is presently no consensus on such an integrity level. However, the likelihood of digital forensics and other forensic techniques to cause wrongful convictions in the criminal legal system clearly constitutes catastrophic failure, and therefore should be held to the highest integrity level, the level where IV&V should be performed independently.

---

[8] IEEE Standard for System, Software, and Hardware Verification and Validation, IEEE Standard 1012-2016, Sept. 2017 (hereinafter referred to as IEEE 1012) (available at https://standards.ieee.org/ieee/1012/5609/).

[9] For example, NASA's Independent Verification and Validation Technical Framework, https://www.nasa.gov/sites/default/files/atoms/files/ivv_09-1_-_ver_p.doc. [Accessed 7 July 2022.

[10] IEEE 1012, p. 196.

The IV&V process must be independent to avoid conflicts of interest that could lead to catastrophic failure. To this end, IEEE 1012 requires technical, managerial, and financial IV&V when testing software and hardware where catastrophic consequences could occasionally occur and where critical consequences will probably occur.[11] Moreover, letting developers certify their own software is a clear conflict of interest, and the IEEE/Association for Computing Machinery Code of Ethics for Software Engineers is clear about the obligation of developers to manage competing aims.[12] Full definitions of technical, managerial, and financial independence from IEEE 1012 are below, but, in brief, the following must all be independent from the group that oversees the design and building of software: personnel, problem formulation, test and analysis tools for IV&V (technical), responsibility for IV&V (managerial), and control of the budget for IV&V (financial).[13]

Specifically, technical independence "[r]equires the IV&V effort to use personnel who are not involved in the development of the system or its elements. The IV&V effort should formulate its own understanding of the problem and how the proposed system is solving the problem."[14] "Technical independence means that the IV&V effort uses or develops its own set of test and analysis tools separate from the developer's tools."[15] And if sharing tools is necessary, "IV&V conducts qualification tests on tools to assure that the common tools do not contain errors that may mask errors in the system being analyzed and tested."[16] This independence requires the exclusion of parties with a stake in the outcome, which for forensic technologies includes forensic labs and law enforcement agencies who, while not financially dependent on developers, have a shared interest in software's acceptance.

Managerial independence "[r]equires that the responsibility for the IV&V effort be vested in an organization separate from the development and program management organizations. Managerial independence also means that the IV&V effort independently selects the segments of the software, hardware, and system to analyze and test, chooses the IV&V techniques, defines the schedule of IV&V activities, and selects the specific technical issues and problems to act on."[17] The IV&V effort must be "allowed to submit to program management the IV&V results, anomalies, and findings without any restrictions (e.g., without requiring prior approval from the development group) or adverse pressures, direct or indirect, from the development group."[18]

Financial independence "[r]equires that control of the IV&V budget be vested in an organization independent of the development organization. This independence prevents situations where the IV&V effort cannot complete its analysis or test or deliver timely results because funds have been diverted or adverse financial pressures or influences have been exerted."[19]

It is clear from these definitions that peer-reviewed publications, while a priceless tool for scientific inquiry, are not a substitute, nor a valid approximation of IV&V when determining reliability or trustworthiness of a deployed system. Peer-reviewed publications form the foundation of scientific advancement, but peer reviewers of scientific publications are not tasked with answering questions like "Should the digital forensics software or results be admissible in court? Is the digital forensics software fit for the evidence in this legal case?" Peer reviewers do not have access to the system itself and are not tasked with assessing its reliability. Peer reviewers are assessing whether a publication deserves the attention of the scientific community, whether the results described deserve the

[11] IV&V with "rigorous" (the highest level) technical, management, and financial independence "is generally required for integrity level 4 (i.e., loss of life, loss of mission, significant social loss, or financial loss) through regulations and standards imposed on the system development" - where "IV&V responsibility is vested in an organization separate from the development organization." IEEE 1012, p. 199.
[12] D. Gotterbarn, K. Miller, and S. Rogerson, "Computer society and ACM approve software engineering code of ethics," Computer, vol. 32, no. 10, pp. 84–88, 1999. doi: 10.1109/MC.1999.796142.
[13] IEEE 1012, p. 198.
[14] IEEE 1012, p. 198.
[15] IEEE 1012, p. 198.
[16] IEEE 1012, p. 198.
[17] IEEE 1012, p. 198.
[18] IEEE 1012, p. 198.
[19] IEEE 1012, p. 198.

attention of other scientists. With respect to specific legal cases, any individual case could go well beyond the bounds of the published studies.

We acknowledge the fact that not every digital forensic technique must undergo peer review, formal testing, or error rate analysis. However, it is unacceptable for any system influencing decisions with potentially catastrophic consequences - especially forensic techniques used in the criminal legal system[20] - to not be managerially, technically, and financially independently verified and validated.

Therefore, we recommend that NIST:

- Expand Sec. 4.8's discussion of validation vs. verification to cover other important distinctions among types of tests, e.g., local vs general testing, quantitative vs qualitative analysis of results.
- State that digital forensics software should undergo IV&V according to its integrity level as defined by IEEE 1012. If NIST does not recommend that digital forensics should be independently verified and validated in accordance with IEEE 1012, it should articulate why digital forensics are an exception to this international practice among professionals.
- Strike KEY TAKEAWAY #2.5 as informal review is an insufficient method for determining the reliability and trustworthiness of high-risk systems like digital forensics. The appropriate method is to use IV&V.

2. Digital forensics should be tested against and governed by standards adhering to principles of due process, openness, consensus, balance, and right of appeal.

We are concerned that this Review understates the limitations of digital forensics. For example, KEY TAKEAWAY #4.7 states that "[e]xtensive tool testing of over 250 widely used digital forensic tools showed that most tools can perform their intended functions with only minor anomalies." The Review does not provide sufficient analysis to justify the statement that "most tools" will only produce "minor anomalies." Especially as the Review explains that "tool testing" is explicitly not validation or verification (p. 49).

Specifically, as it relates to the typical steps in digital forensics investigations in Chapter 4, discussions of the limitations must be expanded. Sec. 4.6 "Identification and Extraction of Artifacts" and accompanying KEY TAKEAWAY #4.2 should be expanded to provide better coverage of limitations and risks of targeted collection. All techniques (e.g., keyword/string search) have inherent limitations (e.g., deriving, e.g., from the limited knowledge of the operator) which are further complicated by data issues (e.g., incomplete or inaccurate metadata values). The limitations can have very important consequences; however, as it is often the case that, once a collection is made, analysts do not return to the uncollected data. Given the limitations and risks, it is important that local validation of the collection techniques used be conducted. This section would benefit from an amplified discussion of these issues.

Section 4.7 Analysis of Results should also be expanded. Once data has been collected, it is crucial to extract the relevant information from the collected data. There are a number of different types of analysis that may be done to extract relevant information, from straightforward information retrieval to network analysis, and so on. Each of these techniques has limitations and so their effectiveness should be locally validated.

We recommend NIST:

[20] See, for example, IEEE-USA "Letter to the National Institute of Standards and Technology (NIST), responding to request for comments on NIST Internal Report 8351-DRAFT *DNA Mixture Interpretation: A NIST Scientific Foundation Review*," November 18, 2021. Available: https://www.nist.gov/document/nist-ai-rfi-ieee001pdf [Accessed July 7, 2022].

- Expand Sections 4.6, 4.7, and 4.9 and KEY TAKEAWAY #4.2 to better address the limitations of digital forensics.
- Strike KEY TAKEAWAY #4.7 and replace it with a TAKEAWAY that better synthesizes the limitations of digital forensics.
- Add to Section 4.9 that the key requirement for at least some types of testing is the ability to draw random samples from specific subsets of a data population, and that not all search tools have the ability to do so.

3. **Understanding the reliability and trustworthiness of digital forensics requires an accounting of all types and sources of error (including random, systematic, human factors, and organizational) and a meaningful characterization of what can go wrong, how likely is that, what are the consequences, and what can be done about it.**

We are concerned with some of the Review's discussion of errors and error rates in Sec. 4.10.

First, there is no evidence that digital processes "tend to have systematic rather than random errors" as conclusory stated in KEY TAKEAWAY #4.4. A digital process may have random errors or systematic errors depending on the type of digital process or digital technique. The article by Lyle referenced in justification of KEY TAKEAWAY #4.4 itself explains that while for some forensic tools like string search, an error rate may be of limited value, for other tools like file recovery and carving, statistical error rates based on random errors can be meaningful.[21] There should be no blanket statement that any specific type of error is automatically more applicable to digital processes or that error rates are only useful for random errors. Errors of any kind, from random to systematic, organizational to human factors, should be considered in-scope unless its limited value is clearly justified by the context and context and type of process being evaluated.[22]

In the context of error rates, the Review should not be so focused on error rates, and instead focus on the key questions of what can go wrong, how likely is that, what are the consequences, and what can and should be done about it.[23] These are the questions central to the fields of reliability engineering and system safety and the theories of probabilistic risk analysis and defense-in-depth. Lyle understood this broader perspective that just because there is not a formal statistical error rate, does not mean there is no meaningful error to be discussed. As he explained, "[t]ools and techniques without a meaningful statistical error rate should have the types of failures and triggering conditions characterized."[24]

Second, it is generally untrue that "[e]rrors in computer science techniques tend to be so small as to be negligible" as alleged in KEY TAKEAWAY #4.5 Such a statement could be interpreted as a license to do nothing and is not acceptable. This KEY TAKEAWAY seems to imply that industry standards in computer science - such as IEEE 1012 - should not apply to digital forensics technology. We strongly disagree with any conclusion of that kind.

We recommend that NIST:

- Expand Sec. 4.10 to use the appropriate reliability engineering and system safety approach to errors which discusses what can go wrong, the odds of something going wrong, the consequences, and what can and should be done about it.

---

[21] J. R. Lyle, "If error rate is such a simple concept, why don't I have one for my forensic tool yet?," *Digital Investigation*, vol. 7, pp. S135–S139, Aug. 2010, doi: 10.1016/j.diin.2010.05.017.

[22] J. H. Saleh, K. B. Marais, E. Bakolas, and R. V. Cowlagi, "Highlights from the literature on accident causation and system safety: Review of major ideas, recent contributions, and challenges," *Reliability Engineering & System Safety*, vol. 95, no. 11, Nov. 2010, doi: 10.1016/j.ress.2010.07.004

[23] Saleh *supra* n. 20.

[24] Lyle *supra* n. 19.

- Rewrite Key Takeaway #4.4 to state "*For some* digital processes, *error rates may be of limited value*. In those cases, an error mitigation analysis provides more information and is the correct way to manage uncertainty. *Tools and techniques without a meaningful statistical error rate should have the types of failures and triggering conditions characterized."* If NIST does not adopt our recommendation, it should include support for this key takeaway, especially as it pertains to the definition of "some digital processes" and how it came to the conclusion that "error rates may be of limited value." These statements appear to be offered without supporting evidence.
- Strike the language in KEY TAKEAWAY #4.5 that "Errors in computer science techniques tend to be so small as to be negligible." If NIST insists on keeping this in, it should explain its supporting evidence.

4. **To ensure digital forensics' effectiveness, competence, awareness, accountability, and transparency in operation, there must be standards and certifications for digital forensics and their operators, and recurring benchmarking exercises and independent studies.**

IEEE believes that trustworthy systems must adhere to principles including effectiveness (system creators and operators shall provide evidence of the effectiveness and fitness for purpose), transparency (that the basis of a particular decision should always be discoverable), accountability (systems shall be created and operated to provide an unambiguous rationale for all decisions made), awareness of misuse (system creators shall guard against all potential misuses and risks in operation), and competence (system creators shall specify and operators shall adhere to the knowledge and skill required for safe and effective operation).[25]

Fully identifying and characterizing the limitations, failure modes and sequences, and error rates is critical to the trustworthy and reliable use of digital forensics. Therefore, rather than concluding in KEY TAKEAWAY #4.6 that it is infeasible to test all combinations of tools and digital evidence sources, IEEE believes the finding should communicate a reasonable standard for testing. While it is true that exhaustive testing is often not possible, we would encourage establishing benchmarks and would like to see NIST encourage and promote standards for reasonable testing and best practices for clear documentation of what testing has been done and what the results were.

We recommend that NIST strike KEY TAKEAWAY #4.6.

Critically, this testing must include the operators and individual forensic laboratories and law enforcement agencies. The Review makes clear each individual operator's significance. We agree with the Review that "[e]ach lab should ensure that personnel understand the basic capabilities and limitations of a tool, especially the relationship between the tool and the fast-changing IT environment." (Sec. 4.8) We also agree with KEY TAKEAWAY #2.4 that **"**[t]he forensic examiner needs to be aware of key changes in computing technology relevant to the examination being performed. Frequent changes in digital technology introduces the possibility for incomplete analysis or for misunderstanding of the meaning of artifacts."

We believe, and NIST should recommend, that governments should make the reports documenting the required IV&V and audits of forensic techniques public. Furthermore, we believe that governments, including NIST, should encourage, develop, and update standards and certifications for digital forensics and their operators, and fund recurring benchmarking exercises and independent studies to ensure their effectiveness, competence, inclusiveness, accountability, and transparency in operation. Specifically, we believe these standards, certifications, exercises, and studies should address:

[25] IEEE Ethically Aligned Design at 18.

- The requirements for informed trust by the general public in digital forensics (see Recommendation #6 below) and the development of metrics that are immediately and easily accessible by experts and non-experts alike;
- The existence or absence of reliable and unbiased underlying scientific principles and methods in digital forensics;
- The requirements for recurring testing and auditing of the operation of digital forensics, including the operators, field conditions, testing data, environments, methodologies, and performance metrics;
- The requirements for publicly available documentation by developers and testers of digital forensics, and of the use of digital forensics in individual and aggregate cases and decisions;
- The requirements for certification or loss of certification of operators and digital forensics, and for their validation for digital forensics already in use;
- The requirements for individuals to be able to access, review, contest, and correct the data about them, to review and contest the decisions that affect them, and to request human review of such data and decisions;
- The requirements for operation in an ethical manner; and,
- The requirements for identifying and addressing vulnerabilities and threats to security, safety, and privacy such as spoofing, evasion attacks, transfer learning attacks, and data poisoning.

5. **Determining the reliability and trustworthiness of forensic technologies like digital forensics requires evaluating them in their operational environments, their use in legal proceedings and how fit the technology was for those uses.**

IV&V is predicated on the value of testing technology in operational environments. No software or hardware is "generally" reliable -- any technology is only fit for certain purposes. Even technologies that are widely considered to be reliable have known failure modes. For example, cellular telephones are widely considered to be reliable but are not classified as "generally" reliable because they do not work effectively in tunnels or underground. Further, the desire for a technology to be classified as "generally " reliable rather than to consider its reliability in a particular case is misguided. A core premise of labeling a product or process as "well-engineered" is that these operating conditions are specifically defined, tested against pre-defined standards, and accompanied with estimated rates of failure. Systems like digital forensics are engineered products incorporating scientific models and therefore require not only the perspective of researchers who have published proofs-of-concept but also engineers who have used product trials and operational testing and evaluation to demonstrate system performance in operating conditions, against predefined standards, and estimated rates of failure.

Therefore, a scientific foundation review of the reliability and trustworthiness of forensic technologies cannot be effective if detached from an analysis of how the technology is used in legal proceedings, in the forensic technology's operational environment -- yet that is exactly what this Review is purporting to do. The Review examines the peer-reviewed and laboratory studies but does not compare that to any of the criminal cases where digital forensics has been used. Notwithstanding the concerns over peer-reviewed studies discussed above, if the types of files, types of encoding, etc. analyzed in legal proceedings are not similar to the samples used in the peer-reviewed or laboratory studies, the studies have little value.

To determine the reliability of digital forensics, we recommend that NIST catalog and evaluate how digital forensics are being used in legal proceedings and how fit the technology is for those uses.

6. **Stakeholders of digital forensics include far more than forensic scientists, attorneys, judges, and juries. They include the public upon whom these systems are used, litigants, academics, journalists, and other researchers. For those users to assess the degree of reliability, validity, and whether that information is fit-for-purpose, they need appropriate access to the software.**

Users are too often inappropriately denied access or forced to overcome improper and unnecessary barriers to access digital forensics in order to determine the degree of reliability, validity, and whether that information is fit-for purpose. There are many more users of digital forensics than merely forensic scientists, judges, or juries. Independent testing of proprietary or government digital forensics by litigants, academics, journalists, and other researchers is needed to ensure that digital forensics are properly vetted and held accountable. NIST should recommend governments clarify whether and how proprietary digital forensics may be reverse engineered, modified, and evaluated under laws such as the Computer Fraud and Abuse Act and the anti-circumvention provision of the Digital Millennium Copyright Act, and rules of procedure and evidence. More broadly, NIST should recommend governments take steps to affirmatively promote awareness, access, research, and testing including:

- Ensuring accountability and transparency in government procurement and contracting for digital forensics;
- Identifying and disclosing the digital forensics used by the government;
- Adopting clear procedures relating to collection, usage, storage and sharing of personal information collected and used by digital forensics;
- Providing constituents notice about digital forensics decisions, explanations for those decisions, and processes for challenging decisions or data; and,
- Specifically, in legal disputes, tribunals should permit disclosure under appropriate protective orders of intellectual property related to digital forensics when necessary to obtain evidence in compliance with other judicial requirements, including constitutional requirements, discovery laws, or subpoenas.

7. Trustworthiness is determined by more than reliability, and therefore, to determine trustworthiness, one must assess the processes and procedures where these systems are deployed.

Technical assessments of reliability as surveyed in the Review are not the sole determination of trustworthiness. There are eight principles for creating and operating systems that further human values and ensure trustworthiness:[26] (i) human rights: systems shall be created and operated to respect, promote, and protect internationally recognized human rights; (ii) well-being: system creators shall adopt increased human well-being as a primary success criterion for development; (iii) data agency: system creators shall empower individuals with the ability to access and securely share their data, to maintain people's capacity to have control over their identity; (iv) effectiveness: system creators and operators shall provide evidence of the effectiveness and fitness for the purpose of systems; (v) transparency: the basis of a particular system decision should always be discoverable; (vi) accountability: systems shall be created and operated to provide an unambiguous rationale for all decisions made; (vii) awareness of misuse: system creators shall guard against all potential misuses and risks of systems in operation; and, (viii) competence: system creators shall specify and operators shall adhere to the knowledge and skill required for safe and effective operation.

Therefore, we recommend that NIST evaluate more than the technical assessments of reliability to determine trustworthiness. Below we list additional requirements for ensuring the trustworthiness of systems in general which includes the automated decision systems such as digital forensics and many of the forensic technologies used today. If those providing or using digital forensics or any other forensic technologies do not adhere to these requirements, then they should not be deemed trustworthy or fit for their use in determining or affecting people's rights and liberties.

To ensure trustworthiness of AI systems, digital forensics, and other forensic technologies, we believe that governments, forensic laboratories, and law enforcement agencies should be required to:

[26] IEEE Ethically Aligned Design.

- Ensure awareness, access, and research on the existence, fairness, safety, security, privacy, and ethical and societal impacts of digital forensics.

Governments should: (i) publicly identify and disclose the digital forensics used by the government; (ii) conduct and publicly disclose a methodological validation study that establishes the value of using new digital forensics in place of existing practices prior to deploying digital forensics; (iii) adopt clear procedures relating to the collection, usage, storage, and sharing of personal information in the context of developing, using, and validating a given digital forensics in a privacy-preserving manner; and (iv) prevent intellectual property, confidentiality claims, lack of funding, or lack of an designated independent body within government to monitor compliance from impeding duly limited independent validation and verification and publicly disclosed review of the fairness, safety, security, privacy, and ethical and societal impacts of digital forensics. Digital forensics ought to be submitted voluntarily to the agency performing validation and verification thereof, and the agency using related private intellectual property or proprietary data in its evaluation must adopt rules to protect such private rights from misappropriation.

Users and the public should be allowed to (i) request and receive an explanation of how a government determination using digital forensics was reached; (ii) determine whether the digital forensics used in government decision-making disproportionately impacts a protected class; and (iii) rectify, challenge, or complete inaccurate or incomplete personal data that is part of the digital forensics system or decision.

- Commit to removing barriers to parties' access to information needed to ascertain relevant evidence about and from digital forensics in legal disputes.

Specifically, in legal disputes where judges, juries, and lawyers are the users of digital forensics results, barriers to parties' access to information needed to ascertain relevant evidence about and from digital forensics should be eliminated.[27] Intellectual property protections should not be used as a shield to prevent duly limited disclosure of information needed to ascertain whether digital forensics meets acceptable standards of effectiveness, fairness, and safety. Specifically, in legal disputes, tribunals should permit disclosure under appropriate protective orders of intellectual property related to digital forensics necessary to obtain evidence in compliance with other judicial requirements, including constitutional requirements, discovery laws, or subpoenas. Furthermore, laws, procedures, and public funding should not make it more difficult for non-government parties in legal disputes to develop, obtain expertise regarding, or gain access to evidence from digital forensics than for government parties to do so.

- Ensure accountability and transparency in procurement and contracting for digital forensics.

To support awareness, access, and research on the existence, fairness, safety, security, privacy, and ethical and societal impacts of digital forensics, there must be accountability and transparency in government procurement and contracting for digital forensics. The government should not procure digital forensics that (i) require the governmental entity to indemnify vendors for any and all negative outcomes; (ii) do not adhere to the eight principles in IEEE's Ethically Aligned Design for creating and operating digital forensics that further human values

[27] For example, when source code is ordered to be provided, "information needed" requires providing sufficient information for the recipient to build, run, and test the software themselves including, at minimum:

- All software dependencies including third-party code libraries, toolboxes, plugins, frameworks, and databases;
- Software engineering and development materials describing the development, deployment, and maintenance of the version(s) of the software system used in the instant case, including software engineering documents and build instructions;
- All records of software glitches, crashes, bugs, or errors encountered during the developmental validation study;
- Software version numbers of the components of the system used for the developmental validation study; and,
- All records of unexpected results, including false inclusions, false exclusions and the conditions under which the unexpected results were achieved.

When source code is ordered to be provided, "access" requires, at minimum, that the source code be made available for inspection, in a format allowing it to be reasonably reviewed, searched, and tested, during normal business hours or at other mutually agreeable and reasonable times, and at mutually agreeable and reasonable locations.

and ensure trustworthiness (as may be reflected in articulated guidelines, standards, certifications, audits, and other sound documentation);[28] (iii) do not comply with federal, state, and local anti-discrimination laws; or, (iv) are shielded from independent validation and verification, and public review.

8. **To ensure digital forensics are reliable and trustworthy, governments should provide sufficient funding for testing, evaluation, certification, and investigation of digital forensics.**

Throughout the document, the Review highlights the value of government funding in the development of research on digital forensics (e.g., the tool specifications, test plans, and data sets for tool testing in Sections 4.10.4 and 4.10.5). We believe NIST should go further by including a KEY TAKEAWAY recommending that governments provide sufficient funding for testing, evaluation, certification, and investigation of digital forensics. The adoption and acceptance of digital forensics requires developing and sustaining public confidence in their quality, reliability, and compliance with regulations and social norms. Increased government funding for government and independent third-party evaluation and certification of digital forensics is essential to ensure efficacy, transparency, traceability, accountability, and competency. Development of design requirements, methods, metrics, and environments so that digital forensics can be tested and evaluated for interactions with different systems is critical in the adoption and acceptance of digital forensics. To this end, mechanisms must be developed for identifying and accounting for the features of digital forensics that could cause current testing, evaluation, certification, and investigation methods to misinform decision makers or the public about the risk of system deployment or the causes of system malfunction.

9. **Take inspiration from NIST's own Text Retrieval Conference (TREC) program.**

It appears that NIST may not be taking advantage of the agency's extant work. NIST's TREC is a strong model that could provide inspiration for the level of rigor in evaluation, measurement, and reporting that NIST seems to be aiming for in this document. While TREC is not directly applicable - the system focuses on the information retrieval phase that comes after the information collection phase - it could provide an example to follow.

[28] IEEE Ethically Aligned Design.