GAO@100

Highlights of GAO-21-435SP, a report to congressional requesters

July 2021

TECHNOLOGY ASSESSMENT

Forensic Technology

Algorithms Strengthen Forensic Analysis, but Several Factors Can Affect Outcomes

Why GAO did this study

For more than a century, law enforcement agencies have examined physical evidence to help identify persons of interest, solve cold cases, and find missing or exploited people. Forensic experts are now also using algorithms to help assess evidence collected in a criminal investigation, potentially improving the speed and objectivity of their investigations.

GAO was asked to conduct a technology assessment on the use of forensic algorithms in law enforcement. In a prior report (GAO-20-479SP), GAO described algorithms used by federal law enforcement agencies and how they work. This report discusses (1) the key performance metrics for assessing latent print, facial recognition, and probabilistic genotyping algorithms; (2) the strengths of these algorithms compared to related forensic methods; (3) challenges affecting their use; and (4) policy options that may help address these challenges.

In conducting this assessment, GAO interviewed federal officials, select non-federal law enforcement agencies and crime laboratories, algorithm vendors, academic researchers, and nonprofit groups; convened an interdisciplinary meeting of 16 experts with assistance from the National Academies of Sciences, Engineering, and Medicine; and reviewed relevant literature. GAO is identifying policy options in this report.

View GAO-21-435SP. For more information, contact Karen L. Howard at (202) 512-6888 or howardk@gao.gov.

What GAO found

Law enforcement agencies primarily use three kinds of forensic algorithms in criminal investigations: latent print, facial recognition, and probabilistic genotyping. Each offers strengths over related, conventional forensic methods, but analysts and investigators also face challenges when using them to assist in criminal investigations.

Latent print algorithms help analysts compare details in a latent print from a crime scene to prints in a database. These algorithms can search larger databases faster and more consistently than an analyst alone. Accuracy is assessed across a variety of influencing factors, including image quality, number of image features (e.g., ridge patterns) identified, and variations in the feature mark-up completed by analysts. GAO identified several limitations and challenges to the use of these algorithms. For example, poor quality latent or known prints can reduce accuracy.

Facial recognition algorithms help analysts extract digital details from an image and compare them to images in a database. These algorithms can search large databases faster and can be more accurate than analysts. The accuracy of these algorithms is assessed across a variety of influencing factors, including image quality, database size, and demographics. GAO identified several challenges to the use of these algorithms. For example, human involvement can introduce errors, and agencies face challenges in testing and procuring the algorithms that are most accurate and that have minimal differences in performance across demographic groups.

Probabilistic genotyping algorithms help analysts evaluate a wider variety of DNA evidence than conventional analysis—including DNA evidence with multiple contributors or partially degraded DNA—and compare such evidence to DNA samples taken from persons of interest. These algorithms provide a numerical measure of the strength of evidence called the likelihood ratio. To assess these algorithms, law enforcement agencies and others test the influence of several factors on the likelihood ratio, including DNA sample quality, amount of DNA in the sample, number of contributors, and ethnicity or familial relationships. GAO identified two challenges to the use of these algorithms. For example, likelihood ratios are complex and there are no standards for interpreting or communicating the results as they relate to probabilistic genotyping.

Generally, three entities test forensic algorithms to ensure they are reliable for law enforcement use.

Source: GAO. | GAO-21-435SP

GAO developed three policy options that could help address challenges related to law enforcement use of forensic algorithms. The policy options identify possible actions by policymakers, which may include Congress, other elected officials, federal agencies, state and local governments, and industry. See below for details of the policy options and relevant opportunities and considerations.

Policy Options to Help Address Challenges with the Use of Forensic Algorithms

Policy option	Opportunities	Considerations
Increased training (report p. 44) Policymakers could support increased training for analysts and investigators.	• Training on human factors could reduce risks associated with analyst error and decision-making. • Could help users or investigators understand and interpret the results they receive. • For latent print and facial recognition, training on cognitive biases could raise awareness and improve objectivity. • Standards for training or certification of analysts or users could increase consistency and reduce risk of improper use across the various federal and non-federal labs and law enforcement agencies that use algorithms.	• Training materials may need to be developed or made more widely available. • May not be clear what entity should establish standards or certifications of training because multiple groups are involved in developing and disseminating training.
Standards and policies on appropriate use (report p. 45) Policymakers could support development and implementation of standards and policies on appropriate use of algorithms.	• Standards or policies addressing the quality of data inputs could reduce improper use. • Increased consistency across law enforcement agencies could increase public confidence. • Standards for testing and performance of facial recognition algorithms could help to reassure the public and other stakeholders that algorithms are providing reliable results. • Standards or policies for communicating results could help users to better understand the strength of the evidence and come to an informed conclusion.	• May be difficult to implement across different levels of government. • Individual localities or agencies may be reluctant to conform to more universal standards. • May increase the cost of procuring and maintaining forensic algorithms. • Standards creation can be resource-intensive, requiring research and testing as well consensus from public- and private-sector stakeholders.
Increased transparency (report p. 46) Policymakers could support increased transparency related to testing, performance, and use of algorithms.	• The public may be more inclined to trust algorithms if officials provide access to the results of testing, and to information about data sources, how algorithms are used, and for what types of investigations. • Increasing the availability of comparative testing results and presenting them in a way that is easy for non-technical users to understand could make it easier for agencies to select the best-performing algorithms. • For facial recognition algorithms, clearly identifying software versions used in testing could improve public confidence and help agencies choose algorithms. • Making more data sets publicly available for facial recognition algorithm training and testing could improve algorithms and minimize demographic effects.	• Algorithm developers may not want to divulge proprietary information related to training and testing. • Sharing the source of training and testing data may create risks to privacy. • Law enforcement agencies or crime labs may have difficulty finding peer-reviewed journals interested in publishing validation studies from testing.

Source: GAO. | GAO-21-435SP