AI, Forensic Identification and the Governance of Criminal Justice

The Economy Research Editorial[1,2]

[1]The Economy Research, 71 Lower Baggot Street, Dublin 2, Co. Dublin, D02 P593, Ireland

[2]Swiss Institute of Artificial Intelligence, Chaltenbodenstrasse 26, 8834 Schindellegi, Schwyz, Switzerland

Abstract

This report considers the use of artificial intelligence in the criminal justice process, focusing on forensic victim identification from skeletal remains, facial images, fingerprints, tattoos, dental records, and other physical evidence. It asserts that although AI is flawed, it should not be prohibited when its ability to reduce search fields, clear backlogs, and hasten identification of unknown victims by using machine-assisted pattern recognition has been established. However, the real problem is an institutional rather than a technical one: the probability outcomes generated by AI may be seen to have the certainty of a DNA result or some other form of scientifically validated forensic corroboration. This report therefore, differentiates AI used in investigation versus AI as evidence. Based on evidence from U.S. Forensic databases and NIST tests, the DOJ and GAO, and China's establishment of judicial AI, the report demonstrates that prohibition rather than judicious regulation is the optimal public policy, allowing AI-assisted identification only within clear parameters of validation, auditability, trained human scrutiny, transparency, external corroboration and specific constraints on its role in the courtroom.

1 Introduction - AI in Criminal Justice and the Forensic Identification Problem

The standard forensic fantasy in Bones is an unusual policy reference. It embodies a quintessential contemporary temptation: that if we can generate identity, a biography and a narrative of crime from a degraded fragment, then perhaps a computer trained on sufficient images, scans and records can do the same thing, faster and cheaper, at scale. It is that intuition that explains the current narrowness of AI debates in criminal justice; these conversations have centered, correctly, on predictive policing, sentencing scores and generative systems' potential for fabrication, but they have largely ignored an adjacent and even more pressing, question: how should criminal justice institutions govern the use of AI when the raw data input is not some wobbly social proxy, but a piece of evidence that appears to have a certain solidness to it – bones, a fingerprint, a face, a tattoo, a dental image, a fading forensic photo? The consensus among reports from the Brookings Institution,[1] the U.S. Department of Justice[2] and Stanford Law School[3] can be distilled to a central point: it is an institutional, rather than a technical, problem. AI is being adopted in criminal justice faster than the agencies that will use it can learn about the tool's mechanics, its failure modes, its civil rights implications and its limits. When these institutions lack the resources for rigorous study of AI, then they must be expected to practice superficial supervision, depend on vendors and fall prey to inflated public trust.

On the narrow issue of using AI for forensic victim identification, this conclusion should strike us deeply. There is nothing intrinsically wrong with using AI to find a victim; machine-assisted pattern recognition in constrained contexts can be a powerful tool and, in specific applications, genuinely brilliant. What is wrong is allowing a lead-generating tool to become quasi-definitive evidence simply because it presents its answers with a statistic or a visually authoritative image. Conventional comparisons to DNA are a misdirection here: DNA profiling isn't merely "accurate technology"-it is part of a mature institutional practice with laboratory standards, chain of custody, validation, probabilistic interpretation and adversarial scrutiny. Many AI-assisted identification tools lack a comparable ecology of reliability; even where the raw material is physical rather than behavioral, it must still be interpreted through a social system: the system that collects and labels it, filters and operationalizes it and decides which errors are acceptable in practice. It is the claim of this report that AI should be treated as a bounded investigative aid for triage, exclusion and hypothesis-formation, not as a self-validating proxy for verified identity, that represents the intended correction to both technological utopianism and outright prohibition. The deeper costs will accrue when this line is crossed, resulting in bias, decrepit investigative practice, asymmetric error burden and a subtle shift in epistemic authority away from trained professionals toward vendors and algorithms.

What we are experiencing is not just a hypothetical now. AI is already beginning to be assimilated into the missing and unidentified persons infrastructure. The National Missing and Unidentified Persons System (NamUs) reported 26,278 active missing-person cases and 15,483 active unidentified-person cases as of February 2026; in fiscal year 2025 to date, it received 93 requests to its image analysis system (which includes anthropologic species analysis, forensic art, facial recognition and tattoo searching) and had 104 cases still pending analysis[4] as of that February 2026 reporting period. FBI NCIC reported 93,447 active missing-person and 8,546 active

unidentified-person records[5] as of the end of 2024.

These are not sideline databases, but part of the core system that has, until now, operated in a context of strained capacity, fragmented reporting and poor data quality. In such a system, AI will not be a neutral productivity increase, but a force that reshapes what cases get looked at, which ones get prioritized, how quickly uncertainty is tolerated and at what point the tentative becomes definitive. And in an uncomfortable but fundamental way, the real question is not whether the technology can do the technical work better than a human-it can and will, in certain applications; rather, it is how institutions will fundamentally recalibrate themselves around that fact. Once the skeletal case gets flagged by image analysis, once candidate identifications are algorithmically ranked, once case backlog gets addressed by probabilistic prioritization and once prosecutors or judges see and react to the result under a veneer of computational precision, human judgment isn't just "in the loop"; it's shaped by the loop. That second-order effect-reviews are less skeptical, vendors sell what the state can't test, best practices atrophy, lawyers lack the technical fluency to challenge the result-is where the system has struggled and, for that reason,Stanford work suggests the criminal justice problem of AI is an institutional one: the problem is that thousands of under-resourced police, prosecution, court and probation departments are utterly unprepared to evaluate the claims of vendors who pitch products at an impossibly technical level.[6] If the state allows AI to help find the dead, it needs to decide, in the first instance, whether it's willing to administer its impact, not just its input.

2 Advocating Voices: The Case for AI-Assisted Forensic Identification

A plain account of why we should use AI in the relevant area should be readily stated, for it is not trivial. Issues of physical identification within the realm of criminal justice typically revolve not around conceptual ambiguity, but issues of scale, degradation and delay: Faces must be searched against large galleries, fingerprints against national databases, skeletal remains must be first typed for species and then reduced to an administratively usable biological profile, cold cases must be reopened across records too large to process by hand.

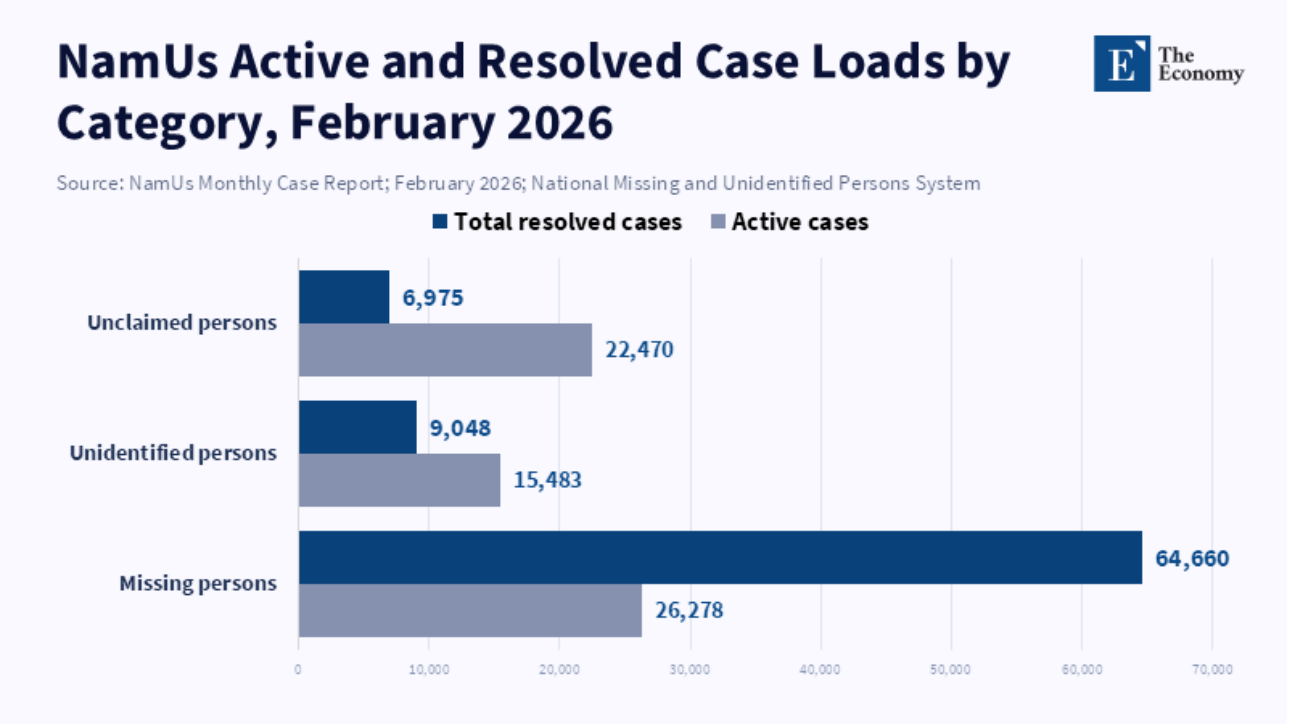

Figure 1: AI-assisted triage becomes attractive where unresolved identification workloads remain large and uneven across case categories.

The 2024 DOJ Report states explicitly that AI may improve the accuracy of identification and forensic analysis over human comparison alone in some contexts[7] and that criminal-justice agencies already rely on biometric AI in order to facilitate facial recognition, fingerprint and associated matching, etc. Its remarks on facial recognition are illuminating: by 2024, contemporary facial-recognition systems routinely returned results with false-negative rates of less than 1 percent at a false positive rate of 3 in 1,000[8] at the 2024 benchmark performance level. The argument of proponents is very straightforward: where there is a large search space and the relevant physical material is image or pattern-like, there is nothing inherently irrational about machine assistance; human comparisons may be slower, more inconsistent and more prone to bias in the unstructured environment of an investigative case.

That argument appears all the more persuasive when we look away from philosophical arguments and towards practice. The National Institute of Standards and Technology has tested almost 200 facial-recognition systems from almost 100 vendors using over 18 million images of over 8 million individuals[9] and continues to perform rigorous tests of extremely large one-to-many identification settings.

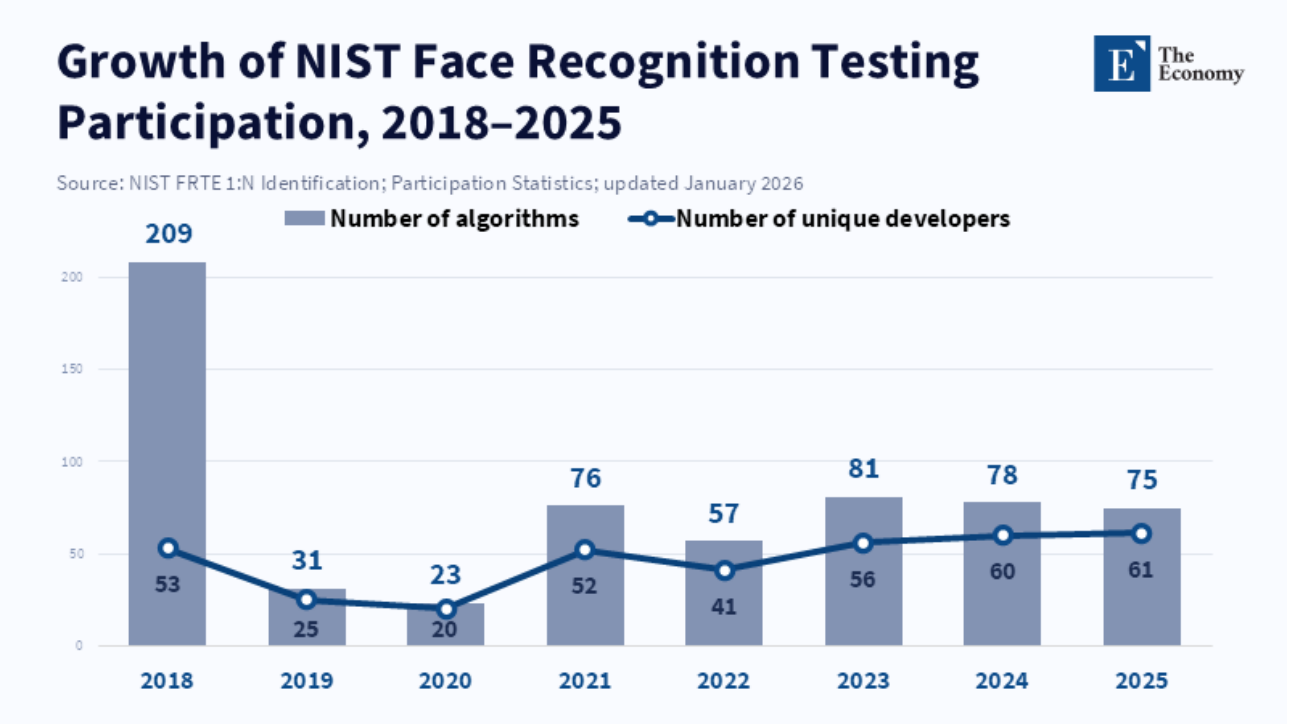

Figure 2: The testing ecosystem has expanded quickly, but broader vendor participation also makes procurement and oversight more complex.

In efforts related to the identification of victims and the analysis of skeletal remains, the National Institute of Justice has sponsored the development of technologies with demonstrated utility as support tools: The system OsteoID was developed to assist in the discrimination of human from nonhuman skeletal remains. NIJ reported that species identification accuracy was 91 percent and human versus nonhuman identification was performed more than 95 percent of the time correctly; some element-specific protocols showed up to 99 percent correct classification[10] if element type was established first.

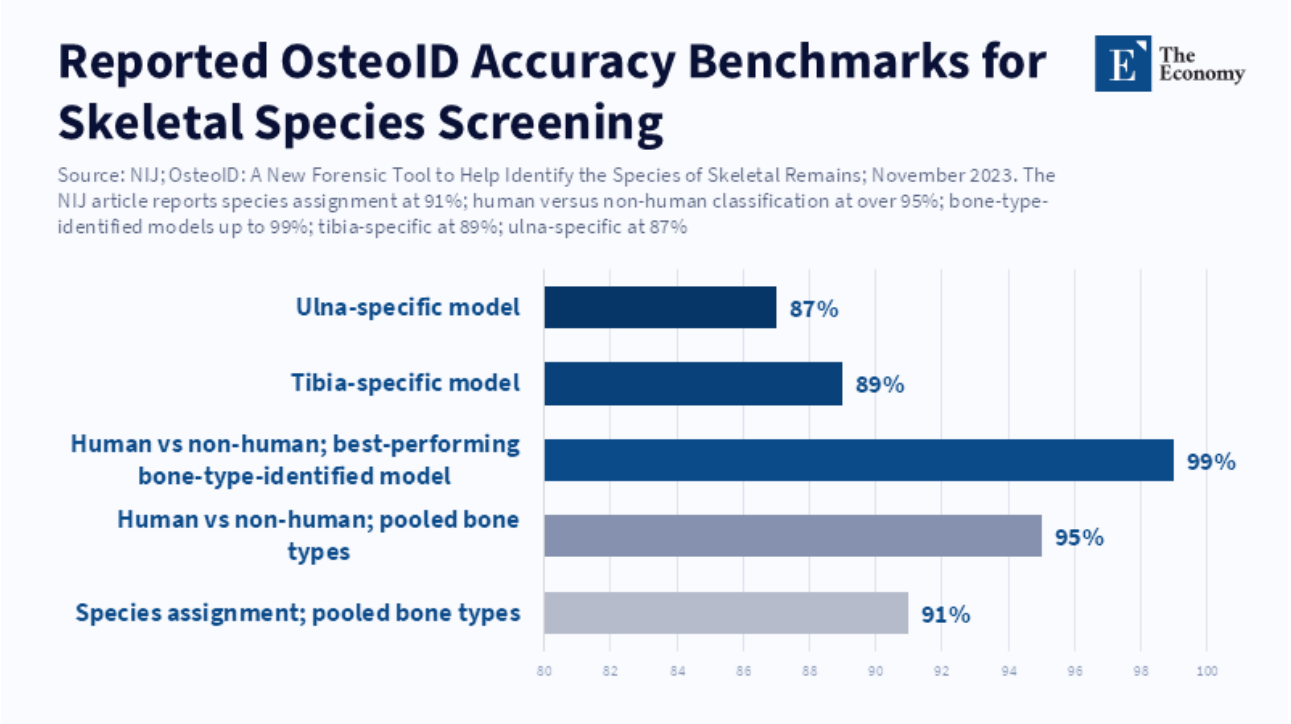

Figure 3: OsteoID shows the value of AI as a screening tool, while also clarifying why screening accuracy should not be mistaken for personal identification.

Meanwhile, a 2025 review of machine learning in forensic anthropology cataloged 167 studies,[11] demonstrating that the use of AI in that discipline has progressed far beyond preliminary inquiry. Biological sex estimation, comparative modeling and skull-based analysis appeared most often, indicating that AI has begun to be utilized as a practical way to narrow the field of possibilities when skeletal material is partial or fragmented. In this way, proponents have identified something correct: the relevant technology is entering an existing and established scientific field rather than an empty one; a discipline already accustomed to using morphological evidence, measurements and cross-modal comparison to re-create or identify an individual from fragments.

However, precisely the evidence that makes the pro-AI case appear credible is also evidence that places limitations on the pro-AI argument: The first limit is one of epistemology. Most of the forensic anthropology literature does not assert the capacity for individual, positive identification from bone alone. It makes statistical estimates of sex, age, stature, proxied ancestry, or species. These are tools that assist in the creation of a search field, rather than definitively settling who was where. The 2025 literature review that cataloged the 167 ML studies of forensic anthropology stressed the continued need for transparency, interpretability and validation[12] of those results. The results of craniofacial identification research remain so variable that a 2024 review emphasized that the standards and understanding of which research findings justify operational implementation remained uneven.[13] In short, the cutting-edge science at present does not yield an AI capable of definitively naming the dead in the way that DNA can. What does exist are tools that may aid in search, prioritization and profile creation under specified conditions, but there is a significant epistemic distance between the two. A helpful narrowing and positive identification are not semantically interchangeable; within a criminal justice system, conflating the two is a clear path towards an investigation based not on evidence, but on suspicion derived from AI outputs.

The second limit is one of operationalization, not concept: Even strong benchmark results reported by NIST fail to fully answer fundamental governance questions. NIST explicitly delineates between automated thresholded identification systems and systems that generate ranked candidate lists for investigative review,[14] particularly in one-to-many settings. Moreover, it is keen to point out that inaccurate identification, whether through false positives or the loss of real matches, carries substantial risks and that its work on demographic-effects indicates performance disparities across different groups: One ongoing NIST evaluation indicated that thresholds designed to maintain a false-positive identification rate of 0.002 among Eastern European females resulted in higher rates in other groups ranging from 20-fold to 80-fold [15] and the exact algorithm varied. These data points do not mean that such systems [16] should be abandoned; they suggest that standard training data sets will not wholly eliminate social risk. Although the input to the system may be a face or bone, the output and downstream consequences rely entirely on decisions about search thresholds, enrollment circumstances and data quality (whether image-based or based on recovered materials). Moreover, in many situations, large databases, degraded imagery and decay phenomena may push system performance well past benchmark tests into the range where an AI result, even if provisional, comes to guide an entire investigation.

The third limit is institutional and may be the most concerning: A common response to machines that perform imperfectly is to suggest ”human in the loop” verification, but the available evidence indicates that this may amount to a placebo rather than a genuine guardrail. Seven DHS and DOJ law enforcement agencies were found by the U.S. Government Accountability Office to be using facial recognition systems without any prerequisite training; reports indicated approximately 60,000 uses of such systems prior to training requirements going into effect, in agencies that provided data to GAO. Far more shockingly, the FBI had just 10 out of 196 employees [17] accessing one such system trained according to the agency’s recommended standard, GAO reviewed. This is the paradigm of automation bias in action: if the human checkers themselves have not been sufficiently trained to recognize what is being input, what a machine’s limits may be, or what specific group failures could be occurring, then their presence only serves to ratify a system operating outside best practice and potentially contributing to errors. It carries the risk of arresting innocents, providing false assurances to families and, most critically, systematically eroding the standards of evidence to the point that a machine-generated inference becomes indistinguishable from a finding.

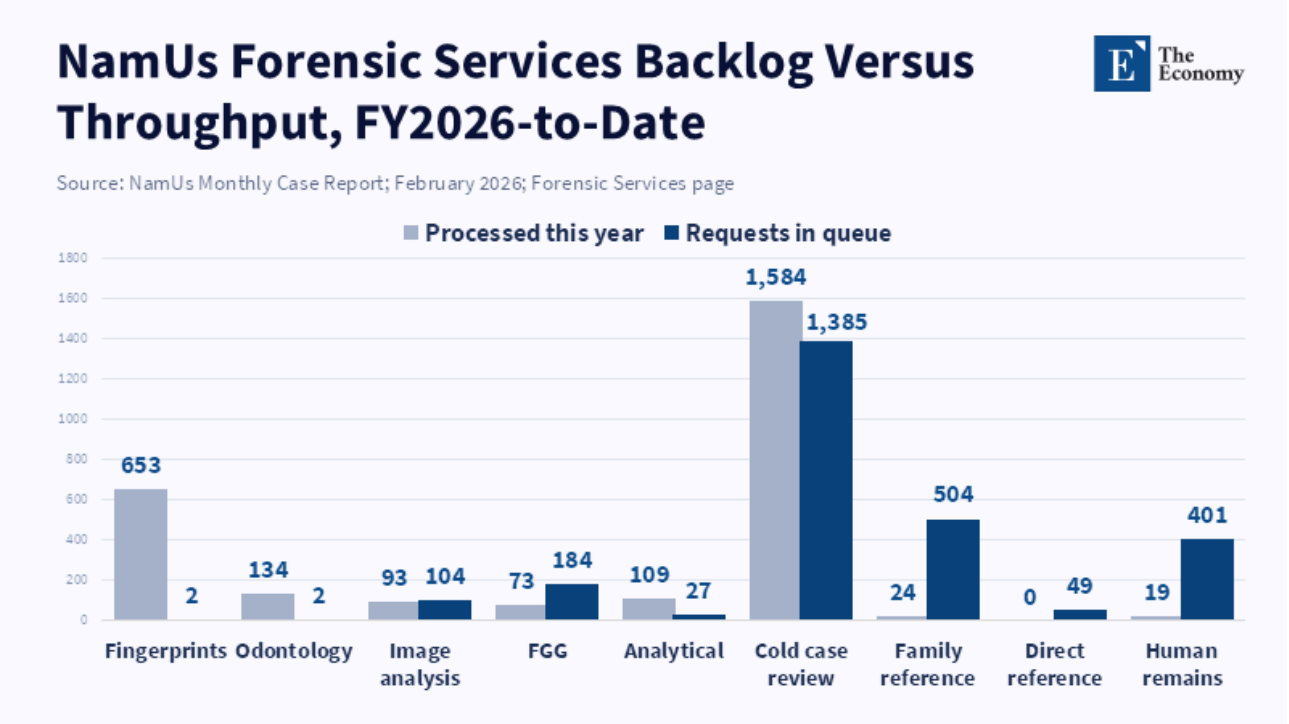

Figure 4: Backlogs are concentrated in the very services where AI-assisted review may be most tempting and most consequential.

For these reasons, the typical comparison to a calculator or search engine is not analytically sufficient. A calculator performs operations whose underlying mathematics are generally clear and errors can be checked relatively easily; a search engine operates on a corpus of documents whose content is transparent. Biometric and forensic AI systems fall into a third category; they may accelerate certain processes, redistribute attention across the analytic landscape and raise the odds that probabilistic evidence will be treated as definitive, owing to the technology's seeming scientific rigor. The case for using AI in these situations, therefore, remains strongest where it remains humble-an efficient tool for expanding searches, reducing processing delays and improving triage. The argument flounders when it attempts to pass off these functionalities as essentially social-risk-free simply because the underlying material does not appear to be sociology. Criminal justice demands an accounting not just of whether a pattern can be detected, but also of the steps involved in transforming that pattern into actionable judgment and in deciding when and how that judgment bears on matters of liberty.

3 How to Govern It: From Investigative Lead to Evidentiary Safeguard

If the central risk is the migration of AI from investigative assistance to de facto proof, the initial governing step must be to draw a sharp legal line between those concepts-a far more critical step than any abstract debate about whether AI should exist in criminal justice. The output of a machine match should be legally designated and treated as an investigative lead unless and until confirmed by a distinct and documented process. When the task is suspect identification, this means no arrest, charge, or detention solely on the output of a facial recognition system or a machine-cued photo lineup. When the task is victim identification of skeletal remains, this means no

machine-generated profile construction, facial approximation, or image matching is to be regarded as definitive personal identification. In all but the most exceptional circumstances, clearly specified confirmation by DNA, fingerprint, dental match, or another validated, multimodal process demonstrably appropriate to the case must be obtained. This is not technology-averse; it's a way of maintaining the distinction between "shrinking the search space" and "providing conclusive certainty of identity." This point of distinction, as the recent Brookings analysis recognizes, is one that states are naturally positioned to define,[18] especially since the potential injuries from unreliable or ill-posed AI applications are often quite real already.

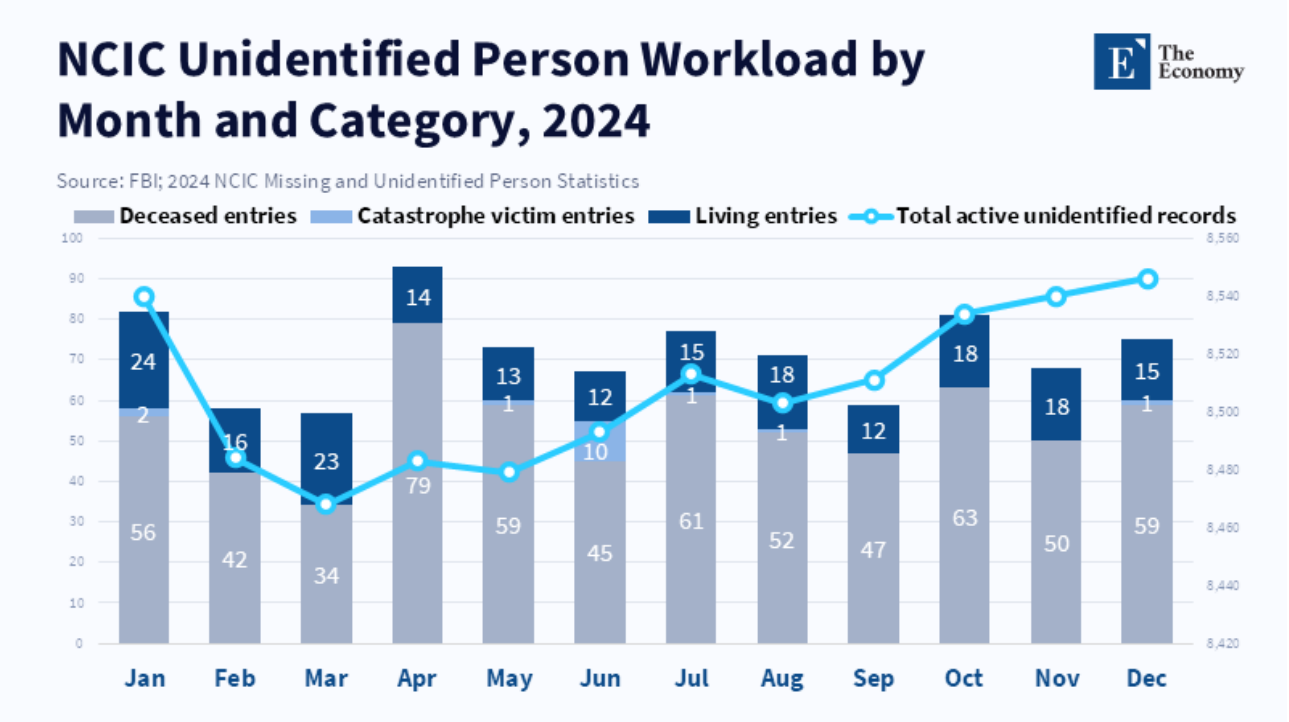

Figure 5: Most unidentified-person entries concern deceased individuals, making forensic identification a high-value but high-risk domain for AI assistance.

That initial step-categorizing the outputs of investigative tools-naturally leads to a second governing task: demanding evidence of system performance aligned with operational practice rather than vendor marketing. The Department of Justice's own internal governance framework now articulates conditions of the kind needed to oversee high-stakes AI use cases. This includes requiring qualitative impact assessment, quantitative testing, continuous performance and bias monitoring and consolidated decision-making before deployment.[19] These are strong starting points. In the specific domain of forensic identification, these principles must be operationalized more sharply. Tools must be tested for a given use case, not merely a product category. A tool that is superb for clean, passport-like photographs might be completely inappropriate for decomposed remains, occluded surveillance camera footage, fragmented tattoos, or databases characterized by inconsistent metadata. Agencies should demand that vendors disclose training data provenance, benchmarking conditions, subgroup error rates, identified failure modes and the extent of performance degradation under actual operational constraints. They should also demand third-party validation before procurement, regular subsequent checks and maintenance of the audit trail, including input data, candidate lists, error rates and user overrides, so that both positive and negative results can be re-examined and challenged. Without this transparency, a state cannot adequately claim

to be differentiating machine assistance from proprietary suggestion.

A third layer of governance must shift focus from technology design to human practice. The conventional slogan ”keep a human in the loop” is too superficial; it neither specifies nor regulates the quality of the human element. For forensic applications, the DOJ itself recognizes that the human review function is necessary not only for spotting potential anomalies but also because if AI-generated analysis becomes evidence in court, the human who conducted the review must be able to attest how the tool was used.[20] This critical insight implies that human reviewers must be trained and credentialed specifically for the class of AI systems they are supervising, that they must understand the inputs the system is designed to process, the truth claims it makes, how to interpret thresholds of significance, variations across demographic groups and the distinction between identifying a pattern and establishing personal identity. Agencies should provide clear guidance on both acceptable and prohibited applications, with explicit mandates to avoid drawing conclusions or taking action when the output of a system reflects undue uncertainty. Defense counsel and the families impacted should be furnished with sufficient transparency regarding underlying data and system use to effectively challenge potential errors. In the context of a criminal investigation, lack of transparency is not simply an administrative inconvenience but a profound procedural injustice. Even the most reliable AI tools can be dangerous if those who use them cannot explain their conclusions.

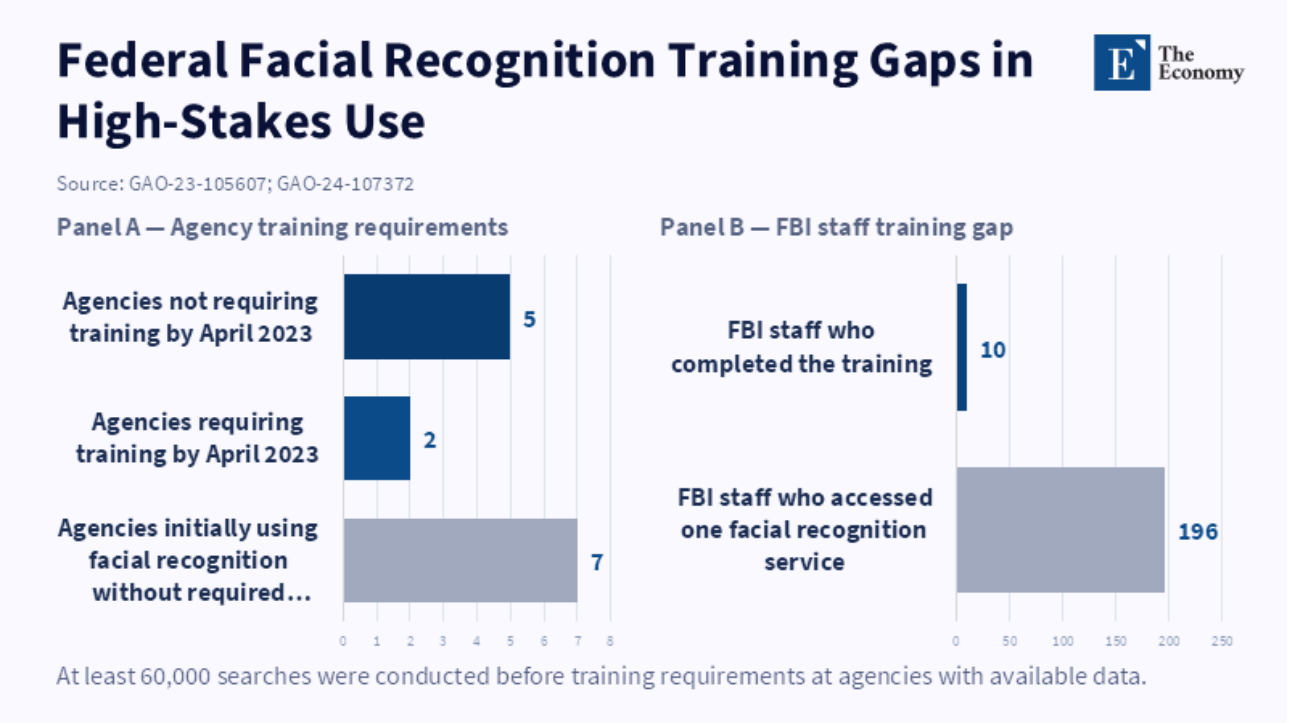

Figure 6: Federal adoption moved ahead of staff preparation, exposing the weakness of relying on human review without specialized training.

A fourth and overarching requirement involves institutional capacity. Here, the Stanford white paper makes a critical point: local criminal justice agencies are often ill-equipped to rigorously evaluate complex AI systems independently,[21] whereas the vendors of these systems are generally sophisticated, well-capitalized entities capable of strongly influencing procurement decisions. Such an asymmetry means that the burden of responsible use cannot be left to individual sheriffs, medical examiners, police chiefs, or trial judges to invent ad hoc policy.

States ought to create dedicated AI governance bodies or commissions for the criminal justice domain, with representation that balances the necessary technical expertise for evaluating systems with legal, ethical and community perspectives that ensure legitimacy. The Stanford working group's proposed design criteria-startup feasibility, relevant expertise, transparency, organizational stability, policy influence and responsiveness,[22] are sensible precisely because they conceive of governance as an ongoing and active process rather than a one-off compliance checklist. They imply processes that enable rapid interim controls when problematic tools are detected and slower, more comprehensive assessments to establish cumulative biases, institutional dependencies, or trends in system performance. The commitment to transparency must also include regular public reporting, open meetings wherever possible, accessible published guidance and an external mechanism through which academic researchers can analyze systems that directly affect people's liberties and identities.

A final layer of governance should also be based on a clear categorization of risk. The least controversial categories involve low-risk applications, such as tools to screen human remains, to de-duplicate existing case files, or to process documents without altering the substantive conclusions. Medium-risk categories include candidate generation in missing persons investigations, image matching between databases, or biologically derived profiling from remains. High-risk uses should include any application whose outputs significantly affect an individual's status regarding arrest, detention, charging, use as trial evidence, or the establishment of personal identity without additional corroboration. Real-time facial recognition of the public for active policing is still high-risk, given its combination of identification with immediate potential for coercive action. Such risk-tiering is necessary because proportionality is also necessary. There are no efficiency gains to be achieved by regulating every application in the same manner, but there is also a grave injustice in allowing agencies to move from low-risk administrative assistance to high-risk judgments of identity without the rigorous standards appropriate to the latter. The OECD's recent review of AI in justice administration offers a pertinent, though indirect, illustration of this distinction: the report notes a variety of tangible efficiency benefits, from case-file management in Spain to accelerated preparation of protective orders in Peru, but also emphasizes potential pitfalls from opacity, overreliance, poor training and insufficient review mechanisms where substantive justice is involved.[23] Streamlined drafting of a document is not synonymous with fair adjudication of a case; quicker identification is not the same thing as justified identification.

The most profound objection against such governance is that it may either obstruct innovation or unduly withhold tools that could help solve cold cases, identify victims more rapidly and reduce investigative burdens. That is a valid concern. There are real opportunity costs associated with delay. The same Department of Justice report that highlights wrongful arrests also acknowledges that the appropriate use of AI-driven identification and forensic analysis can excel human comparative performance and unlock analytical capabilities beyond manual techniques.[24] However, this precisely underscores the need for governance that is ex ante rather than reactive. In an institution as high-stakes as criminal justice, the cost of delay in developing robust protocols can exceed the cost of initially slow adoption. Once an agency begins to reorganize its investigative processes around machine triage, prosecutors' caseloads come to include AI-shaped case files, or families have been convinced that an AI-powered reconstruction has reliably identified their loved ones, the social and emotional costs of reversal are far greater. The goal of governance is not to suppress competent tools but to require that competence be

implemented under conditions commensurate with the potential risks. The policy question is not whether AI should be employed in identifying the dead. It is about how much evidentiary humility a state is willing to maintain when a tool begins to prove too compellingly successful.

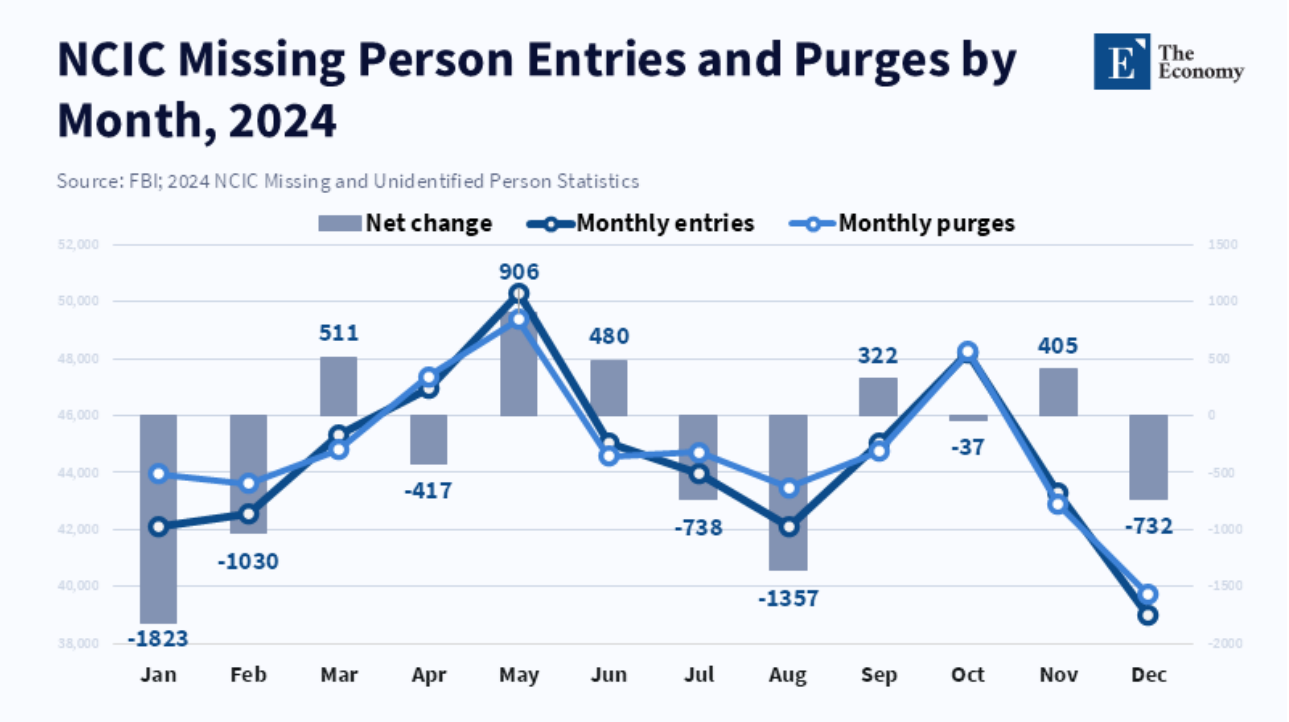

Figure 7: High monthly turnover does not eliminate system pressure; it explains why speed-enhancing tools can reshape investigative priorities.

4 Chinese Case: AI Justice Infrastructure and the Risk of Systemic Overreach

The contemporary experience of China most clearly shows, on a massive scale, both the promise and the peril of using AI as criminal justice infrastructure rather than discrete software. According to the 2025 Oxford Institute of Technology and Justice atlas, China's criminal-justice AI ecosystem is nationwide and encompasses law enforcement, procuratorates, courts and defense. Tools listed included the 206 System, Sharp Eyes, Skynet, the National Judicial AI Platform and numerous integrated support systems.[25] The tasks set for these tools, correspondingly, range widely from administrative support, case processing and charging support to data analysis, decision-making support, legal drafting support, operational support and predictive analysis. Training, notes the same atlas, "is not mandatory or systematic" and while a Supreme People's Court directive for AI in criminal proceedings was issued in 2022 and general rules apply to criminal procedure and data protection, there is "no specific legislation on AI in criminal proceedings". This is a telling configuration: widespread deployment, broad functional ambitions and patchy procedural specificity. In institutional terms, this implies an assumption that integration should precede governance, that structure can be imposed layer by layer.

A 2025 HSE article on AI in Chinese criminal justice offers additional context, situating it as a top-down "judicial intelligence movement" spurred by national policy since the 2017 New Generation Artificial Intelligence Development Plan.[26] It states that courts and procuratorates all over China developed their own AI-powered

judicial platforms and that use has spread to the overwhelming majority of regions. The article, balanced in its approach, concedes that AI may optimize resources and enhance judicial efficiency, crime prevention and investigation, while also flagging "familiar risks" such as lack of standards, fragmented data, disparities in data availability [27] and issues of legality, decision accuracy and fairness. It goes on to express concern that AI may incrementally narrow judicial discretion even where it doesn't formally displace judges and proposes an "inclusive regulation model" combining technological, legal and ethical components-a clear signal that technical efficacy alone cannot solve legitimacy challenges once the tools are framing the decision points for charging, detention and prosecution sentencing support.

The significance of the Chinese case is that it so plainly demonstrates the speed with which throughput gains can themselves come to constitute public narratives of legitimacy. The Oxford Atlas, for example, reported that the Hainan High People's Court saw a 50 percent increase in production speed and a 70 percent decrease in the time it takes to produce written judgments after adopting AI, with drafting of procedural documents falling by 90 percent.[28] Such numbers are hard to ignore and explain why advocates for judicial AI in China and abroad emphasize potential relief for overburdened systems. Yet these very numbers reveal the limits of efficiency as a measure of system health: a speedier writing process alone doesn't assure us whether parties can challenge the process, whether low-training staff grasps its limitations, whether burdens fall disproportionately, or whether decision-makers are implicitly tuning their reasoning to a system's available outputs. In essence, quantifiable time-saving doesn't prove robust adversarial fairness, reliable evidentiary integrity, or true human accountability-it merely proves faster writing and the inability to discern between these two outcomes constitutes systemic failure.

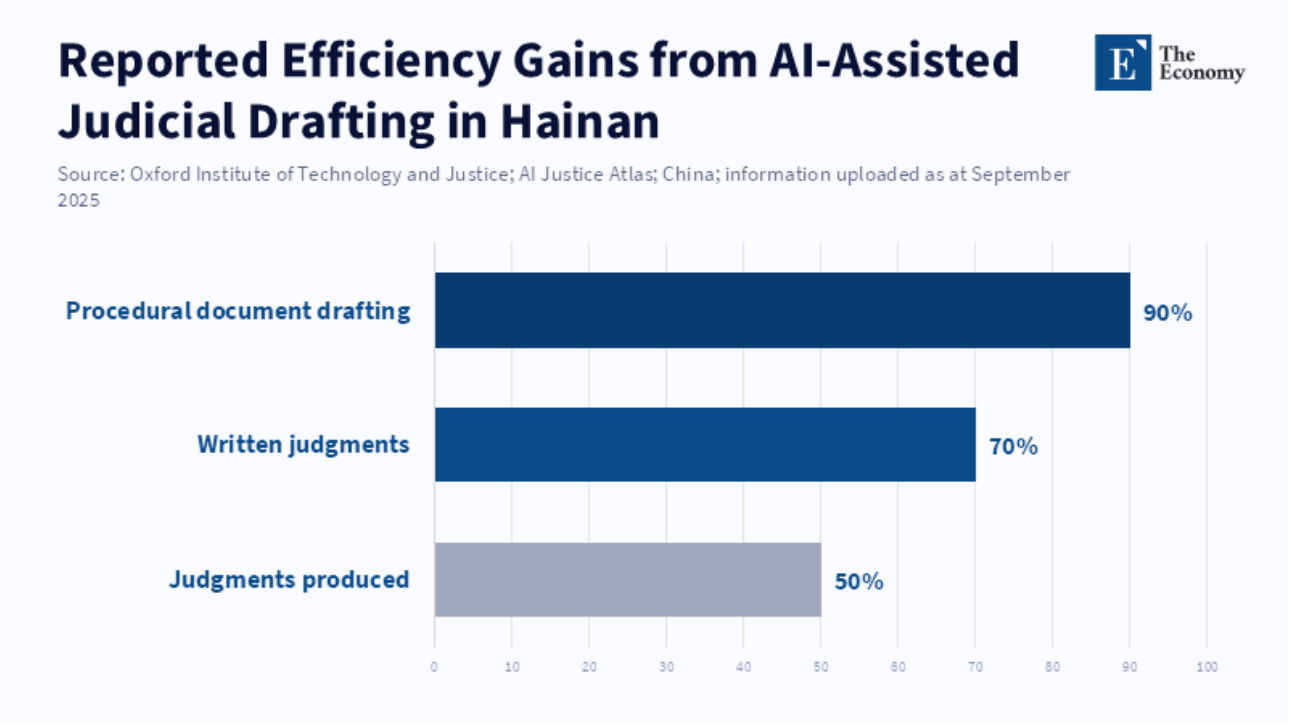

Figure 8: China's reported gains are strongest in administrative speed, but faster processing does not by itself prove evidentiary reliability or procedural fairness.

For forensic victim identification specifically, the implications of the Chinese example are uniquely stark. The

model appears to suggest the rapid merging of disparate forensic tools-identification systems, case-management technologies, surveillance cameras and decision-support platforms-into an undifferentiated administrative apparatus. The conceptual distance between finding the victim, identifying the perpetrator and preparing the prosecution file is diminished; in a fully integrated system, the process may appear seamless but could also lead to an undocumented chain of inferences. One ambiguous result might shape investigative priority, which in turn might shape what evidence is presented and prioritized and subsequently, what conclusions prosecutors draw and what approach judges take, all without ever definitively confirming the initial uncertainty. It is the latter structural risk, one often missing from simpler Western debates, that is most alarming. China's judicial AI does not demonstrate the impossibility of justice through technology per se, but something far more nuanced and pressing: once embedded within workflow, a technology's limited accountability makes for a system that loses reversibility, forfeits transparency and requires a significant increase in institutional self-restraint-precisely the kind of restraints that cannot exist if victim identification AI is governed by procurement paperwork alone, but instead requires meaningful law, robust disclosure obligations and impartial oversight.

This suggests, initially, that the central premise may have been only half correct. AI can indeed allow criminal justice institutions to take on certain tasks, once the domain of painstakingly slow human analysis: spotting trends, identifying fragments, generating candidates for recognition and serving as the ultimate identifier for those whose selves are lost. With judicious use, this may preserve both accuracy and efficiency; assist in the fundamental task of differentiating human from non-human remains; revive cold cases; or shorten the agonizing period of the unknown. Yet AI also tempts such systems into reducing investigative possibilities to the certainty of evidence. The problem isn't that physical evidence is too subjective for machine analysis, but that criminal justice systems so easily accept machine analysis of physical evidence as having inherent certainty beyond its own capacity. Recent evidence points in both directions: benchmark progress is real;[29] deployment is already pervasive;[30] some lower-stakes uses are clearly helpful; and yet validation is erratic, training is weak,[31] demographic biases persist[32] and the tools have already been misused in ways that threaten liberty. The path forward is obvious: states and justice agencies should use AI-assisted victim identification only in a regulated framework that mandates its outputs be understood as tentative leads, requires independent corroboration before making any final decision, necessitates ongoing verification and audits and establishes enduring public institutions tasked with overseeing its evolution over time. The core question is not whether AI can identify the victim; it is whether systems using this technology can maintain the humility of evidentiary reasoning or whether technology-driven certitude will supplant the rule of law.

5 Conclusion - Keeping AI Assistance Below the Threshold of Proof

The very same claim most supportive of AI in criminal justice is one that most obviously demonstrates its insufficiency. Machine systems performing forensic identification, in circumstances involving bones, poor pictures, or shattered physical evidence, can reduce the work, increase filtering speed and lighten the administrative burden, which allows victims to go unnamed for years. That function is valid. But it does not grant permission to equate probabilistic pattern-recognition with biological fact. The fundamental policy error lies not in deploying

the machine, but in the subtle elevation of its evidentiary weight: a classification tool may be prompted, by institutional inertia, to perform conclusion-forming beyond its epistemic capacity.

That risk seems acute precisely where data sets are most seemingly definitive. Bone comparison, image-matching and biometrics might appear more objective than other forms of pattern prediction, but are fundamentally shaped by training data, the models chosen, thresholds determined and human interpretation. In that context, errors are distributive, legal and irreparable. A misplaced identification may misdirect an inquiry, delay naming a deceased body, or place a free individual under the specter of compelled interrogation.

The right policy intervention here is not complete prohibition nor complete embracing. It is controlled governance-validated use, defined evidentiary space, mandated human supervision, transparent logging and clear adherence to the rule that AI can inform forensic practice, but it can never perform substitution for confirmation of identity when it involves freedom and name.

References

[1, 18] Bains, C., Chohlas-Wood, A. and Kinsey, K. (2026) 'States can—and should—regulate AI in criminal justice', *Brookings Institution*, 16 April.

[2, 7, 20, 24] U.S. Department of Justice (2024) *Artificial Intelligence and Criminal Justice: Final Report*. Washington, DC: U.S. Department of Justice.

[3, 6, 21, 22] Stanford Law School (2026) 'AI in criminal justice: why governance matters and how to make it work', *Stanford Law School*, 27 March.

[4] National Missing and Unidentified Persons System (2026) *NamUs Monthly Case Report: February 2026*. Washington, DC: National Institute of Justice.

[5] Federal Bureau of Investigation (2025) *2024 NCIC Missing and Unidentified Person Statistics*. Washington, DC: FBI Criminal Justice Information Services Division.

[8, 14, 29] National Institute of Standards and Technology (2026) *Face Recognition Technology Evaluation: 1:N Identification*. Gaithersburg, MD: NIST.

[9, 15, 32] National Institute of Standards and Technology (2019) *Face Recognition Vendor Test Part 3: Demographic Effects*. NISTIR 8280. Gaithersburg, MD: NIST.

[10] National Institute of Justice (2023) 'OsteoID: a new forensic tool to help identify the species of skeletal remains', *NIJ Journal*, 15 November.

[11, 12] Faisal, E. and Rogers, T.L. (2025) 'A review of the literature on the applications of machine learning in forensic anthropology', *Forensic Science International*, 376, 112579.

[13] Wilkinson, C., Liu, C.Y.J., Shrimpton, S. and Greenway, E. (2024) 'Craniofacial identification standards: a review of reliability, reproducibility, and implementation', *Forensic Science International*, 359, 111993.

[16] U.S. Government Accountability Office (2023) *Facial Recognition Services: Federal Law Enforcement Agencies Should Better Assess Privacy and Other Risks*. GAO-23-105607. Washington, DC: GAO.

[17, 31] U.S. Government Accountability Office (2024) *Facial Recognition Technology: Federal Agencies' Use and Related Training*. GAO-24-107372. Washington, DC: GAO.

[19] U.S. Department of Justice (2024) *Department of Justice Compliance Plan for OMB Memorandum M-24-10*. Washington, DC: U.S. Department of Justice.

[23] OECD (2025) *Governing with Artificial Intelligence: The State of Play and Way Forward in Core Government Functions*. Paris: OECD Publishing.

[25, 27, 30] Oxford Institute of Technology and Justice (2025) *AI Justice Atlas: China*. Oxford: Blavatnik School of Government, University of Oxford.

[26, 28] Higher School of Economics (2025) 'Artificial intelligence in Chinese criminal justice', *Law in the Digital Age*, 2025.