Passagens. Revista Internacional de História Política e Cultura Jurídica
Rio de Janeiro: vol. 18, nº 2, maio-agosto, 2026, p. 294-310.

https://doi.org/10.15175/tx5x0930

Article

The algorithmic mind: Artificial Intelligence and the transformation of forensic psychology

Swarnima GORANI*
Rahul J. NIKAM**
Sharda University, Greater Noida, Uttar Pradesh, India

Abstract

This paper explores the intersection of Artificial Intelligence (AI) and forensic psychology and how AI technologies are transforming the traditional space of forensic psychological testing, criminal profiling, risk assessment, and trial testimony. Drawing on current usages, integrating empirical research conducted, and critically evaluating the resulting ethical and legal implications, this paper reveals that, despite the fact that AI provides levels of analytical power and pattern-recognition skills never before offered, it also creates considerable problems in the form of bias, indirectness, and the fundamental nature of psychological knowledge. The study argues that the future of forensic psychology will not be one that includes the wholesale replacement of human expertise by AI, but it will be one that integrates it considerately so that professional judgment is not compromised but rather utilized to take advantage of technological evolutions. The research presents an in-depth evaluation of how AI is able to transform the field of forensic psychology, and it provides a future outlook on how to implement AI in criminal justice-related scenarios.

Keywords: AI & law; forensic psychology; behavioral sciences; algorithm; law.

A mente algorítmica: Inteligência Artificial e a transformação da psicologia forense

Resumo

Este artigo explora a interseção entre Inteligência Artificial (IA) e psicologia forense, bem como a forma como as tecnologias de IA estão transformando o campo tradicional dos testes psicológicos forenses, da elaboração de perfis criminais, da avaliação de risco e dos depoimentos em julgamentos. Com base nos usos atuais, integrando pesquisas empíricas realizadas e avaliando criticamente as implicações éticas e legais resultantes, este artigo revela que, apesar de a IA oferecer níveis de poder analítico e habilidades de reconhecimento de padrões nunca antes vistos, ela também gera problemas consideráveis na forma de viés, indireção e na natureza fundamental do conhecimento psicológico. O estudo argumenta que o futuro da psicologia forense não será aquele que inclui a substituição total da expertise humana pela IA, mas sim aquele que a integra de forma ponderada, de modo que o julgamento profissional não seja comprometido, mas sim utilizado para aproveitar as evoluções tecnológicas. A pesquisa apresenta uma avaliação aprofundada de como a IA é capaz de transformar o campo da psicologia forense e oferece uma perspectiva futura sobre como implementar a IA em cenários relacionados à justiça criminal.

Palavras-chave: IA & direito; psicologia forense; ciências comportamentais; algoritmo; direito.

---

* Ms Swarnima Gorani is an Assistant Professor at Sharda School of Law, Sharda University, Greater Noida, India. She is pursuing her PhD at Himachal Pradesh National Law University, Shimla, in Forensic Psychology. E-mail: goraniswarnima@outlook.com. https://orcid.org/0009-0004-2052-6891

** LLM Degree in 2008 and a Ph.D. degree in 2012 from the NALSAR University of Law, Hyderabad, India. He is currently a Professor at Sharda School of Law, Sharda University, Uttar Pradesh, India. E-mail: rahulsnikam@gmail.com. https://orcid.org/0000-0001-7279-1399

Received on January 27, 2026 and approved for publication on May 9, 2026.

Este trabalho está licenciado com uma Licença Creative Commons Attribution 4.0 International License.

La mente algorítmica: la inteligencia artificial y la transformación de la psicología forense

Resumen

Este artículo analiza la intersección entre la inteligencia artificial (IA) y la psicología forense, y cómo las tecnologías de IA están transformando el ámbito tradicional de las pruebas psicológicas forenses, la elaboración de perfiles criminales, la evaluación de riesgos y el testimonio en los juicios. Basándose en los usos actuales, integrando la investigación empírica realizada y evaluando críticamente las implicaciones éticas y legales resultantes, este artículo revela que, a pesar de que la IA proporciona niveles de capacidad analítica y habilidades de reconocimiento de patrones nunca antes ofrecidos, también crea problemas considerables en forma de sesgos, indirectas y la naturaleza fundamental del conocimiento psicológico. El estudio sostiene que el futuro de la psicología forense no consistirá en la sustitución total de la experiencia humana por la IA, sino en una integración prudente de esta, de modo que el juicio profesional no se vea comprometido, sino que se utilice para aprovechar las evoluciones tecnológicas. La investigación presenta una evaluación en profundidad de cómo la IA es capaz de transformar el campo de la psicología forense y ofrece una perspectiva de futuro sobre cómo implementar la IA en escenarios relacionados con la justicia penal.

Palabras clave: IA & derecho; psicología forense; ciencias del comportamiento; algoritmo; derecho.

L'esprit algorithmique : l'intelligence artificielle et la transformation de la psychologie légale

Résumé

Cet article explore les liens entre l'intelligence artificielle (IA) et la psychologie légale, et examine comment les technologies d'IA transforment les domaines traditionnels que sont les tests psychologiques légaux, le profilage criminel, l'évaluation des risques et les témoignages devant les tribunaux. En s'appuyant sur les utilisations actuelles, en intégrant les recherches empiriques menées et en évaluant de manière critique les implications éthiques et juridiques qui en découlent, cet article révèle que, bien que l'IA offre des capacités d'analyse et des compétences en reconnaissance de formes sans précédent, elle soulève également des problèmes considérables liés aux biais, au caractère indirect de ses conclusions et à la nature fondamentale des connaissances psychologiques. L'étude soutient que l'avenir de la psychologie légale ne consistera pas en un remplacement total de l'expertise humaine par l'IA, mais en une intégration réfléchie de celle-ci, de sorte que le jugement professionnel ne soit pas compromis, mais plutôt mis à profit pour tirer parti des évolutions technologiques. La recherche présente une évaluation approfondie de la manière dont l'IA est capable de transformer le domaine de la psychologie légale, et elle offre une perspective d'avenir sur la manière de mettre en œuvre l'IA dans des scénarios liés à la justice pénale.

Mots-clés : IA & droit ; psychologie légale ; sciences du comportement ; algorithme ; droit.

算法思维：人工智能与法律心理学的变革

摘要

本文探讨了人工智能 (AI) 与法律心理学 (Forensic Psychology) 的交集，以及人工智能技术如何改变法庭心理测试、犯罪行为记录、风险评估和庭审证词等传统领域。作者借鉴了当前的AI应用，整合了已有的实证研究，并批判性评估了AI的司法运用以及由此产生的伦理和法律影响。研究表明，尽管人工智能提供了前所未有的案件分析能力和犯罪识别技能，但它也带来了诸多问题，例如偏见、间接性以及心理学知识的基本问题。研究认为，法律心理学的未来在于：人工智能并不能完全取代人类专业知识，司法界应该以一种审慎的方式将AI融入其中，从而确保专业判断不受影响，并能充分利用技术进步带来的优势。作者深入评估了人工智能将会如何改变法律心理学及其相关领域，并展望了它在刑法各类场景中的应用前景。

关键词：人工智能与法律；法律心理学；行为科学；算法；法律

Der algorithmische Verstand: Künstliche Intelligenz und der Wandel der forensischen Psychologie

Zusammenfassung

Dieser Beitrag untersucht die Schnittstelle zwischen künstlicher Intelligenz (KI) und forensischer Psychologie und beleuchtet, wie KI-Technologien den traditionellen Bereich der forensisch-psychologischen Testverfahren, der Täterprofilierung, der Risikobewertung und der Zeugenaussagen vor Gericht verändern. Auf der Grundlage aktueller Anwendungsbeispiele, unter Einbeziehung durchgeführter empirischer Forschung und durch eine kritische Bewertung der sich daraus ergebenden ethischen und rechtlichen Implikationen zeigt dieser Beitrag, dass KI zwar ein bisher unerreichtes Maß an analytischer Leistungsfähigkeit und Mustererkennungsfähigkeiten bietet, jedoch auch erhebliche Probleme in Form von Verzerrungen, Indirektheit und der grundlegenden Natur psychologischen Wissens mit sich bringt. Die Studie argumentiert, dass die Zukunft der forensischen Psychologie nicht in einer vollständigen Ersetzung menschlicher Expertise durch KI bestehen wird, sondern in einer umsichtigen Integration, sodass das fachliche Urteilsvermögen nicht beeinträchtigt, sondern genutzt wird, um die Vorteile technologischer Entwicklungen zu nutzen. Die Untersuchung bietet eine eingehende Bewertung dessen, wie KI das Feld der forensischen Psychologie verändern kann, und liefert einen Ausblick darauf, wie KI in strafrechtlichen Szenarien eingesetzt werden kann.

Schlüsselwörter: KI & Recht; forensische Psychologie; Verhaltenswissenschaften; Algorithmus; Recht.

Introduction

Forensic psychology emerged as a distinct discipline at the intersection of psychology and law, evolving significantly over the past century. The field's roots can be traced to the late 19th and early 20th centuries, when psychologists first began applying their expertise to legal matters. In 1893, James McKeen Cattell conducted pioneering research on the psychology of testimony, examining the accuracy of eyewitness accounts. Shortly afterwards, in 1908, Hugo Münsterberg (2007) published *On the Witness Stand*, which explored how psychological principles could inform legal proceedings, though this works initially met skepticism from the legal community. The early decades saw gradual acceptance as psychologists contributed insights into witness reliability, jury behavior, and criminal profiling.

The mid-20th century marked professionalization and expansion for forensic psychology. World War II accelerated developments in psychological assessment that soon found applications in criminal justice. The 1950s and 1960s witnessed increased collaboration between psychologists and courts, particularly in evaluating criminal responsibility and competency. From the 1970s onward, the field experienced substantial growth with the establishment of the American Psychology-Law Society (1969) and the American Board of Forensic Psychology (1978), formalizing professional standards. Research expanded into child custody evaluations, risk assessment, and false confessions. The rise of DNA evidence in the 1980s-1990s highlighted the fallibility of eyewitness testimony, prompting renewed research into memory and perception. Contemporary forensic psychology has become a sophisticated, research-driven field with numerous specializations.

The domain of forensic psychology now exists at a major crossroad where law is subject to examination, and science is subject to investigation. Modern practitioners conduct court evaluations, consult on jury selection, assess violence risk, and work within correctional facilities, emphasizing evidence-based practices while addressing ongoing challenges in bias reduction and ethical complexities. The years of clinical judgment, empirically validated protocols, and commonplace/ standardized testing have allowed practitioners to answer crucial questions about defendant competence, offender recidivism potential, and voluntary confessions and witness credibility (Heilbrun, 2001). These decisions directly influence the decisions of verdicts and sentencing, as well as the freedoms of people trapped by the criminal justice system.

The introduction of AI in forensic psychology is a remarkable opportunity and a significant challenge. Machine-learning algorithms have the ability to identify patterns in millions of case files that cannot be identified by human experts, and natural-language-processing methods may recognize deception indicators with explicit precision. Predictive models can estimate the recidivism rates by considering hundreds of variables at once. However, the same technologies give extremely alarming signals in terms of bias in algorithms, the lack of transparency of multifaceted models, and the degree of effectiveness of machine cognition in actually understanding the interplay of human behavior and mental states in a complex way. However, the use of AI in forensic psychology is the equivalent of analyzing the introduction of standardized psychological testing in the first half of the twentieth century, and there should be both benefits and drawbacks in its implementation when the lives of human beings are at stake.

Legal Material and Methodology

The current study uses a systematic literature review approach that will reportedly combine both empirical studies, theoretical studies, and current case law and professional guidelines up to 2024. The research explores AI uses in various areas of forensics, such as risk analysis, competency analysis, credibility analysis, criminal profiling, and courtroom testimony, relying upon peer-reviewed scholarly materials, as well as court records, to provide an interdisciplinary approach to AI as a field. AI here is used in a broader sense to refer to machine learning, natural language processing, neural networks, and other computational systems that enable algorithms to accomplish tasks previously performed by human intelligence. Forensic psychology refers to the use of psychological science with regard to a legal question and context, including criminal, civil, and family-law cases. This question is

relevant to current arguments about technology in criminal justice because it focuses on the psychological aspect, in particular, a facet that has traditionally been less discussed than algorithmic sentencing or predictive policing, but also has an equivalent consequence.

Results and Discussion

Forensic psychology is a crucial area that bridges the gap between law and psychology. The study of forensic psychology means applying the scientific principles to the justice delivery system. When this domain is touched by the AI, it is susceptible to AI improvement and to AI weaknesses as well. Knowing this sphere would shed light on grander questions about human judgment, technological potential, and the values by which they should be combined.

Traditional forensic assessment is a synthesis of various streams of data, which may include clinical interviews, collateral information, psychological testing, and record reviews. The principles of forensic assessment emphasize that valid conclusions must presuppose the collection of multimethod data, a knowledge of the style of responding, and context (Heilbrun, 2001). The idea behind this multimethod approach is that human behavior is complex and that single indicators are unreliable. This has been subjected to extensive academic questioning to establish the reliability and validity of forensic assessment. The influential meta-analysis showed that statistical prediction is typically equivalent to clinical prediction, with some key caveats about the quality of both methods. This, alongside other works, set the empirical basis of the implementation of actuarial techniques - the forebearers of modern AI in forensic practice.

The shift towards actuarial assessment began to be taken seriously in the 1990s, including the Violence Risk Appraisal Guide (VRAG) and the Static-99 of sexual-offender risk (Russell; Norvig, 2020). These instruments rely on statistically developed formulas based on empirically validated aspects to compute the risk score. These actuarial approaches are transparent, consistent and can be proved to make predictions as opposed to clinical judgment, which lacks such predictive ability. However, actuarial instruments are very limited. They are standardized on particular populations and cannot be generalized across different situations or demographics. They instead favor capturing static historical variables (e.g., criminal history, age at first offence) instead of dynamic variables (e.g., current mental state, treatment progress). Most importantly, they provide group-level probabilities that might not be accurate when applied to individual cases.

This discussion has always been framed by the debate between actuarial and clinical approaches, which traditionally promotes the argument of mechanical versus holistic prediction. The experts argue that straightforward actuaries invariably beat professional clinical judgment. Critics claim that clinical expertise allows taking into account certain special factors and localities that are not present in formulaic approaches. This conflict sets the scope of arguments about AI in the modern context. Modern AI involves a number of technologies that are interconnected. Machine learning will allow systems to polish operations by means of experience without direct programming. Deep learning utilizes neural networks that are multi-layered networks. Natural-language processing allows human language to be computational. Computer vision promotes the interpretation of images and videos (Russell; Norvig, 2020)

The distinction between modern AI and previous actuarial tools is that it is capable of identifying patterns of that complexity that cannot plausibly be defined by human experts. Machine-learning algorithms can automatically identify predictive regularities in large datasets as opposed to requiring researchers to define the variables that predict what. This feature is not only the biggest strength of AI but also the origin of the most serious concerns and is explored in the course of this paper. Many systematic reviews have covered the use of AI in a forensic context. The analysis of the criminal justice of the algorithm system recognizes risk assessment as the most deep-rooted AI usage in the field. The analysis conducted by Skeem and Lowenkamp (2016) indicates that the use of risk-assessment tools is being adopted by a good number of people who have not been thoroughly validated across most contexts. The systematic review of violence-risk prediction also indicates that although AI methods have promise, they currently do not potently outperform rigorously-verified actuarial tools.

The proliferation of AI to novel fields, such as detecting deception through linguistic analyses, predicting judicial rulings, and computerized grading of psychological tests has marked the dawn of a new era. However, according to the controversial study conducted, recidivism risk prediction made by AI did not show any benefits over its counterparts that were conducted by non-experts in the criminal justice field, which emphasizes underlying concerns about whether algorithmic sophistication can result in better performance. (Dressel; Farid, 2018). There is a growing literature that aids in gaining a better understanding that evaluation must also be conducted with regard to dimensions of fairness, transparency, and procedural justice; predictive accuracy is solely a subset of this. This expanded evaluative paradigm recognizes that in forensic situations, the way decisions are made is carried by the forensic decisions.

The most common types of forensic assessments include a clinical interview, testing instruments, like the MacArthur Competency Assessment Tool-Criminal Adjudication (MacCAT-CA), and expert opinion development. The recent studies discussed the skill of assessing competence using AI. A study was conducted which resulted in predicting evaluator decisions using demographic, clinical and legal variables based on evaluations of over 27,000 patients resulted in the model being able to make predictions with an average accuracy of 89 per cent, indicating that AI may be used to identify cases in which competency can be easily established compared to cases in which the individual needs an in-depth evaluation.

Nevertheless, there are still notable anxieties about AI-enhanced competency evaluation. Legal competency is ultimately a subjective assessment, which considers both the cognitive skills and how well the defendant has an understanding of a specific charge and whether he or she is able to work with a specific attorney. The competency assessment requires ongoing observation of the developing understanding of the defendant, which makes it impossible to convert it to a fixed forecast (Roesch; Zapf; Hart, 2010). Existing AI applications are suitable for use to screen and prioritize instead of replacing full clinical assessment.

Malingering, deception, or exaggeration of symptoms to gain an outside advantage is a significant complication in forensic evaluation. The conventional method of detection is based on the use of validity scales integrated into the questionnaire in the form of the Minnesota Multiphasic Personality Inventory-2 (MMPI-2) and special instruments, like the Structured Interview of Reported Symptoms (SIRS). AI provides new points of entry to malingering. It is provided that the machine-learning analysis of patterns of responses in MMPI-2 can reflect the occurrence of weak inconsistency, which is a symptom of invalid responding (Ben-Porath; Tellegen, 2008). Symptom narratives can be analyzed through linguistic analysis relying on natural-language processing to detect untypical language patterns that do not match the reported impairments. Such studies as reaction-time and eye-tracking employing AI can demonstrate the cognitive effort, Not In line with the claims.

In a comparative study, it was found that AI-generated linguistic patterns on malingered and true PTSD accounts, and found an 87 per cent hit rate in distinguishing faithful and false tasks, far outperforming clinicians on similar tasks. The AI singled out idiosyncratic linguistic features, such as excessive use of extreme language, lack of sensory description, and temporal atypists. Although such promising results have been achieved, AI-based malingering detection is ethically questionable. The phenomenon of over-detection can deprive real symptoms, especially in groups that manifest in culturally different ways. Adversarial contexts of forensics provide defendants with valid reasons why they should be

careful, and AI systems, which are intended to detect deception, can misinterpret such carefulness as a fake. One of the most demanding aspects of forensic work is the retro-evaluation of the mental conditions of a defendant during the commission. The burden of evaluation obliges the reviewer to recreate the mental condition on the basis of crime-scene evidence, witness reports, confessions made by the defendant, and historical documents, none of which provides direct access to the psychological experience in the past. The following are some notable areas where the use of AI in the arena of Forensic Psychology is currently being experimented with, some of which are discussed herewith.

Mental State at Time of Offence (MSO)

The use of AI in mental state of the offender (MSO) assessment is currently still in its exploration phase, but it promises. Pattern analysis can compare the characteristics of the offender's behavior to databases of crimes performed by individuals with a specific mental state. Pattern analysis may make comparisons between the behavioral characteristics of the offender and databases of crimes performed by offenders who display a particular mental state. The researchers at the Institute of Psychiatry, Psychology and Neuroscience of King’s College, London have applied an AI system to examine the records of homicide cases and forecast the occurrence of successful insanity defenses.

It used more than 600 variables, such as the nature of the victim, the type of weapon used, the location, the prior criminal history of the defendant, and the behavior at the crime scene. Its accuracy rate was 76 per cent, similar to that of the highly-trained forensic psychiatrists. However, MSO assessment is an illustration of the oracularities of AI to capture subjective experience. Algorithms may be able to pick recognizable patterns that correlate with a certain mental state, but not how such a state was actually experienced by a particular person at a certain time (Buchanan, 2000).

Psychological Autopsy

There are specific issues with psychological autopsy, a victim-oriented retrospective study of the psychological state of a deceased person, which is most often used in the case of equivocal deaths. Ancestral methodology deals with the interview of informants, records, and the re-creation of the life and psychological environment of the dead in the period before death. The use of AI software can include the analysis of digital footprints, including social media posts, internet searches, and communications, to find indicators of psychological

distress (Coppersmith *et al*., 2018). Photographs can be evaluated using computer vision techniques to determine visual cues of depression or substance use.

In a study, the Twitter posts of individuals who went on to die by suicide were analyzed using natural language processing methods to identify linguistic features, such as more self-referencing, hopelessness, and death-related talk that differentiated pre-suicidal posts by these individuals compared to regular people. This model then resulted in an area under the curve (AUC) of 0.87, which was a good discriminative ability. AI application in psychological autopsy poses a deep worry about privacy, despite its prospects. The act of analyzing digital communications of a deceased person in a way that violates their consent is obtrusive even when the analysis is permitted under the law, and algorithms can misread context - e.g., sarcasm, literary allusion, interests of fiction can be considered a false ideation (Coppersmith *et al*., 2018).

Neuropsychological Assessment

Neuropsychological testing in the forensic setting can respond to the impact of brain damage on functioning and criminal responsibility. The conventional testing involves the use of standard testing of memory, attention, and executive functioning, among other areas of cognition. Neuropsychological assessment is changing due to AI, as the examiners and scoring are automated, which minimizes the variability of examiners and enhances efficiency. Data mining of test scores using machine learning is also capable of identifying a subtle cognitive signature unique to a given state (Heilbronner *et al*., 2009).

An AI system has been created to perform neuropsychological test batteries to forecast functional outcomes in patients with traumatic brain injury. This approach performed better than conventional clinical prediction, especially on complex cases with multifaceted comorbidities. The algorithm found non-linear relationships among clinical deficits, which often remain unnoticed by clinicians. However, automated assessment is not able to reflect the qualitative observations that are of central importance to neuropsychological assessment, such as the way examinees deal with tasks, deal with frustration, or know how to compensate for the deficit. These observations of processes often provide more understanding of the real-world operation than test scores do. While exploring the proliferation of an AI tool known as Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) for neuropsychological assessment, a controversy resulted, which is notable and is discussed here.

Case Study- The COMPAS Controversy

The COMPAS system has caused more controversy than any other AI application in the field of forensic psychology to date. It is a machine learning system that Northpointe (now Equivant) developed to predict the likelihood of recidivism based on 137 variables, such as criminal history, substance abuse, employment, and social environment (Brennan et al., 2009). ProPublica reported on COMPAS in 2016, alleging racism against black defendants after an investigation involving over 7,000 defendants in Broward County, Florida, revealed that black defendants were nearly twice as likely as white defendants to be falsely identified as being high-risk (45s versus 24s), and white defendants were more likely to be falsely identified as being low-risk (48s versus 28s). The prosecution decided that COMPAS was discriminatory towards blacks. Northpointe challenged such results by stating that COMPAS passed a different fairness test, calibration, which states that criminal predictions aligned with the recorded recidivism rates in racial groups (Dieterich; Mendoza; Brennan, 2016).

This brought forth a spirited fight concerning the relative significance of various metrics of fairness. When the base rates of recidivism vary between them mathematically, it is not possible to meet all desirable fairness criteria. The COMPAS scandal exemplifies some of the most basic issues in AI risk analysis. The historical patterns that algorithms possess when they are trained on historical data will replicate the past trends, including the systemic bias in these trends (Selbst; Barocas, 2018). To the extent that historically, Black defendants have been subjected to increased surveillance, more arrests when the same behavior is exhibited, and more severe penalties, the algorithm will emulate historical trends to determine that Black defendants are more prone to risky behavior.

The accuracy of COMPAS and other similar tools has been studied several times. A widely cited study by Dressel and Farid found that COMPAS predicted recidivism with only 65 per cent accuracy, which is just slightly more accurate than chance and no more accurate than the predictions of untrained laypeople with little information accessible to them. Further studies provide a more detailed evaluation. The analysis of a variety of risk assessment tools revealed that none of the tools is presently that effective at the individual level; still, they can be much more effective than chance in identifying risk groups at the group level (Skeem; Lowenkamp, 2016).

Notably, the predictions compared by human clinicians are not perfect and biased as well. Professional judgment in a structured form, that is the combination of actuarial scores and clinical override, can enhance the accuracy, yet it opens up possibilities of bias.

Challenges and Implications

The Black Box Problem

The black box problem is perhaps the most essential problem: most AI systems, especially deep learning models, do not tell how they arrive at conclusions. They find out intricate trends about high-dimensional data that are not accessible to human interpretation. Such secrecy is incompatible with criminal justice transparency standards. Suppose a deep learning model that predicts the risk of violence with 80 0 per cent without errors. When it categorizes a defendant as high risk, we are able to tell that the algorithm detected patterns within the associated data related to violence.

Nevertheless, we are unable to define what patterns, in what way they interplay, and why such a combination is risky to this or that individual in particular (Selbst; Barocas, 2018). The defendant has no meaningful right to contest a prediction that is not explainable. According to some, we take black boxes as a default practice; we do not know how our smartphone or car functions, and we do trust them. Criminal justice, however, is different. The right to comprehend and question the evidence presented against a person is essential when his/her liberty is involved. Accuracy is not sufficient, and the process should be fair, and the fairness means explainability.

Algorithms Discrimination

AI systems also learn based on historical information that inherits historical biases. For instance, in such a case, when historical data is more severe on Black defendants, including being arrested more often for the same behaviors, charged more often with more serious offences, and sentenced harshly, the algorithms will be trained on this information and learn that Black defendants are higher risk. The algorithm does not find evidence of differential criminality; it reinforces differential treatment (Selbst; Barocas, 2018).

Ethical Issues

The integration of artificial intelligence into forensic psychology presents profound ethical challenges threatening justice and fairness. Machine learning systems trained on historical data often perpetuate societal biases related to race, socioeconomic status, and gender. Risk assessment tools like COMPAS have demonstrated racial disparities, with

Black defendants frequently assigned higher recidivism scores than similarly situated white defendants. The opacity of AI systems creates the "black box" problem, where forensic psychologists cannot understand how algorithms reach conclusions, undermining due process rights. Delegating clinical judgment to algorithms risks dehumanizing evaluations and overlooking contextual factors that human psychologists would consider. Questions of accountability arise when predictions prove incorrect, and data privacy concerns emerge as AI requires vast personal information without adequate consent. Many practitioners lack training to critically evaluate AI tools or recognize when algorithmic recommendations conflict with clinical judgment. The field must establish ethical frameworks that prioritise human oversight, algorithmic transparency, bias testing, and protection of individual rights.

Rise in the number of cases

A number of cases have been taken to appellate courts about the usage of AI in criminal justice. In the case of State v. Loomis, 2016, the question was raised of whether the application of the COMPAS risk scores in the sentencing process amounted to a violation of due process. Although the Wisconsin Supreme Court upheld the use of COMPAS with restrictions, the U.S. Supreme Court refused to hear the case, leaving the constitutional issues unanswered. According to the Loomis decision, there were a number of issues regarding COMPAS: it relied on data in another jurisdiction (national data used to evaluate Wisconsin defendants), it was proprietary, and it used gender in a way that might have violated equal protection.

However, the court stated that these considerations did not furnish sufficient reasons to preclude it because COMPAS was not exclusive but one of the factors. This is one of many factors; limitation is an important but possibly illusory one. The studies on psychological anchoring have shown that when a figure is brought up, even explicitly, when it is defined as non-determinative, it still affects later judgment. The risk scores might affect judges who are instructed not to use them alone, yet they could be affected unconsciously. In Malenchik v. State, 2010, the court discussed the application of proprietary psychological testing software. The court decided that the defendant had the right to have his expert investigate the algorithm to determine its legitimacy in cross-examination, which created a contradiction with trade secret defenses by technology enterprises.

Jury Comprehension

Regardless of the fact that AI evidence can be admitted, the jurors are supposed to comprehend what it entails and what its limitations are. Studies in juror understanding of probabilistic evidence are depressing-jurors often misunderstand statistical evidence, overweight a lot of evidence and ignore other evidence that is relevant and are not able to contain judgments conditional on probabilities. The predictions of AI are probabilistic by their nature: the risk of 70 per cent of recidivism means that there is a chance of not reoffending 30 per cent, but this specific defendant still may belong to either category (Skeem; Lowenkamp, 2016). Jurors can ignore a high-risk rating as a virtual certainty where it is only a high probability based on group data. Jurors may overrate scientific evidence due to an exaggerated expectation moderated by the CSI effect, which could extend to AI. Jurors can adopt an assumption that algorithmic conclusions are objective and infallible when such conclusions are actually probabilistic, and an assumption is made. This is a greater concern when the AI evidence is seen to benefit the prosecution since jurors tend to give it an unreasonable weight.

Expert Testimony and Applications in the Courtroom.

To affect legal decisions, AI-based forensic psychology must meet evidentiary admissibility standards. The U.S Supreme Court in Daubert v. Merrell Dow Pharmaceuticals, Inc., 1993, provided guidelines about scientific evidence and stated that it had to be testable, peer-reviewed, error-bound, generally acceptable, and it had to be designed to be used in litigation, as opposed to scientific purposes. There are specific issues of admissibility in relation to AI applications. The problem of the black box, i.e. the impossibility of describing the work of complex algorithms arriving at conclusions, contradicts the requirements of understanding expert testimony and exposing it to cross-examination. Moreover, AI systems are often proprietary, which does not allow an independent validation. The source code of COMPAS, including that, is a trade secret. This does not allow the defendants to explore the possibility that the particular algorithm used in their case is legitimate and impartial.

In the Indian context, the admissibility of forensic psychological tests in Indian courts is governed by the Bharatiya Sakshya Adhiniyam, 2023, which permits expert psychological opinion under Section 45. The landmark Supreme Court judgment in Selvi v. State of Karnataka (India), 2010) fundamentally transformed forensic psychological testing by ruling that narcoanalysis, polygraph tests, and brain mapping cannot be conducted without explicit consent, as involuntary administration violates Article 20(3) of the Constitution of India

protecting against self-incrimination. Results from these tests cannot be treated as confessions but may provide corroborative evidence. Traditional assessments, including intelligence testing, personality evaluations, and competency assessments, face fewer restrictions but must meet reliability and relevance standards. Courts evaluate expert testimony based on the qualifications and scientific validity of the expert. Sections 22 and 23 of the Bharatiya Nyaya Sanhita, 2023 allow mental health opinions regarding criminal responsibility due to unsoundness of mind or intoxication. Challenges include the absence of standardized protocols, limited training programs, cultural adaptation issues with Western tests, and insufficient judicial understanding, requiring clearer evidentiary standards and legislative frameworks under the new criminal justice system.

However, there is no known case where the testimony generated in Forensic Psychology through the usage of AI has been admitted by the court of law in any proceedings.

Conclusion and Future Outlook

The concept of artificial intelligence is changing forensic psychology in some irredeemable ways. Infrastructure technologies discussed herein, such as risk assessment algorithms, detection systems, automated testing of psychology, and predictive profiling, provide what just a few years ago appeared to be features of science fiction but now exist in the marketplace. AI is able to examine trends on scales of millions of cases, discover small interactions that would be invisible to the human eye, and predict with a specific level of accuracy. These possibilities will make forensic assessment more regular, evidence-based, and productive. But this is a promise that is accompanied by grave dangers.

The historical biases can become possible and even more sought after through the same algorithms that are supposed to spread objectivity. Procedural justice and human dignity may be compromised in relationships created to serve the purpose of enhancing accuracy. Efficiency-increasing technologies can lessen transparency and accountability. The black-box challenge implies that we cannot, and do not tend to be able to, explain algorithmic inferences, which makes the very right of people to know and challenge evidence difficult. In the propensity to technological solutionism, that human behavior is large-scale, the degree of complex, irresolvable problems possessing algorithmic solutions, is apt to the disrespectful concretism of the behavior in its complexity. The aforementioned discussion has shown that AI is neither a panacea nor a poison in forensic psychology, but a potent tool whose effectiveness solely relies on the thoughtfulness of its use. There are a number of major conclusions which so appear:

Firstly, AI needs to enhance and not substitute human judgment in forensic psychology. Algorithms are also good in pattern recognition with the large data sets, but do not replicate the contextual insights, moral reasoning, and empathic interaction that define expert psychological evaluation. The best solution is to incorporate human wisdom and responsibility with the power of AI to analyze. Second, transparency and explainability are not a privilege. When freedom is to be considered, people have a right to know which evidence is brought against them. The models within the forensic applications must put more emphasis on the fact that can be understood in an interpretable way, rather than marginal accuracy based on opaque systems. Complex algorithms require a lot of testing and ex post justification.

Third, algorithmic fairness involves continued close monitoring. Unless programmed to fight them, AI systems trained on historical data will recreate historical biases. Controlling bias requires more than technical solutions and requires a fundamental attitude to issues of justice, equity, and what values criminal legal proceedings should embody. Disparate impact should be audited frequently. Fourth, professional norms should develop. This is because it is the bureau of responsibility of forensic psychologists to know the AI tools they are using, prove them to be involved in a particular use and a given population and in any way, make independent judgments and not blindly follow the algorithmic results. To guarantee the use of AI, professional organizations should identify guidelines that help uphold the use of AI and establish compliance mechanisms.

Fifth, procedural justice is just as significant as accuracy. Although the predictions offered by AI might be entirely accurate, it would not preclude the fact that the application of such technology would cast doubts regarding human dignity and autonomy as well as individualized consideration. Since the decision-making process highly influences the validity of the legal system and the adherence to decisions made by the system. These procedural issues and substantive accuracy must also be addressed once the AI implementation occurs. Sixth, it needs regulation and oversight. The market forces will not guarantee that AI systems are justified, reasonable, and well-applied. A mixture of professional self-control, judicial review and possible legislative measures is needed to avoid early adoption of inadequately validated systems or even the inappropriate use of even judged validated systems.

In the future, the future of AI in forensic psychology will be informed by decisions that the profession makes currently. Alternatively, we may engage in unquestioning adoption and put into practice all types of algorithmic aids that boast of efficiency without concerning transparency or fairness. Or we can also accept reactionary rejection, not involving AI and

missing out on improvement. Or we can pursue a middle course: critically incorporating AI where it proves to really work better, demanding openness and justice, preserving human autonomy and liability, and remaining prepared to review further as experience mounts. The gambling was no less than life and death. The direct influence of forensic psychology is in incarceration, child custody, involuntary commitment and the credibility of the testimony. The correct determination of these is really a question of justice. Artificial intelligence will not make predictions perfect human behavior is too complicated and contingent to do so. However, used wisely, it could make forensic evaluation a bit more precise and uniform, as well as define its own flaws more clearly.

The profession of forensic psychology is at a crossroads. It is the way ahead that does not involve either the oblivious idealism or automatic opposition, but an incisive inquiry based on factual wisdom and moral self-examination and uncompromising fidelity to equity. As these technologies keep flourishing, in all matters that challenge us to pose hard questions, such as accuracy, fairness, transparency, dignity, and the nature of the criminal legal regime we desire to be, shall depend our readiness to ask hard questions of AI will make it the most treasured or its most concerning burden. The psychological mind cannot be substituted by the algorithmic one, but it can be aided by the latter. It is an ideological approach that must be applied to all AI applications in the context of forensic psychology: technology in the service of justice rather than technology as an adjunct to judgment. It is only through keeping this distinction that it is possible to harness the promise of AI and avoid the dangers, so that forensic psychology can remain faithful to its core purpose, such that it will be able to apply psychological science to legal issues in a way that is both beneficial to the advancement of the truth and fairness.

How to cite this article:

ABNT

GORANI, Swarnima; NIKAM, Rahul J. The Algorithmic Mind: Artificial Intelligence and the Transformation of Forensic Psychology. *Passagens rev. int. hist. pol. e cult. jur.*, Niterói, v. 18, n. 2, p. 294-310, May.-Aug. 2026. https://doi.org/10.15175/tx5x0930

APA

Gorani, S., & Nikam, R. J. (2026). The Algorithmic Mind: Artificial Intelligence and the Transformation of Forensic Psychology. *Passagens rev. int. hist. pol. e cult. jur.*, *18*(2), 294-310. https://doi.org/10.15175/tx5x0930

Copyright:

Copyright © 2026 Gorani, S., & Nikam, R. J. This is an Open Access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original article is properly cited.

Publisher responsible for the evaluation process:

Gizlene Neder

References

BEN-PORATH, Yossef S.; TELLEGEN, Auke. *MMPI-2-RF*: manual for administration, scoring, and interpretation. Minneapolis: University of Minnesota Press, 2008.

BUCHANAN, Alec. *Psychiatric aspects of justification, excuse and mitigation in Anglo-American criminal law*. London: Jessica Kingsley Publishers, 2000.

COPPERSMITH, Glen *et al*. Natural language processing of social media as screening for suicide risk. *Biomedical Informatics Insights*, v. 10, p. 1-11, 2018. https://doi.org/10.1177/1178222618792860

DIETERICH, William; MENDOZA, Christina; BRENNAN, Tim. COMPAS risk scales: demonstrating accuracy, equity, and predictive parity. Northpointe Inc., 8 July 2016.

DRESSEL, Julia; FARID, Hany. The accuracy, fairness, and limits of predicting recidivism. *Science Advances*, v. 4, n. 1, eaao5580, 2018. https://doi.org/10.1126/sciadv.aao5580

HEILBRONNER, Robert L. *et al*. American academy of clinical neuropsychology consensus conference statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist*, v. 23, n. 7, p. 1093-1129, 2009. https://doi.org/10.1080/13854040903155063

HEILBRUN, Kirk. *Principles of forensic mental health assessment*. New York: Springer, 2001.

MÜNSTERBERG, Hugo. *On the witness stand*: essays on psychology and crime. [*S. l.*]: Kessinger Publishing, LLC, 2007.

ROESCH, Ronald; ZAPF, Patricia A.; HART, Stephen D. *Forensic psychology and law*. [S.l.]: John Wiley & Sons, 2010.

RUSSELL, Stuart; NORVIG, Peter. *Artificial intelligence*: a modern approach. 4th ed. Hoboken, NJ: Pearson, 2020.

SELBST, Andrew D.; BAROCAS, Solon. The intuitive appeal of explainable machines. *Fordham Law Review*, v. 87, p. 1085-1139, 2018. https://dx.doi.org/10.2139/ssrn.3126971

SKEEM, Jennifer L.; LOWENKAMP, Christopher T. Risk, race, and recidivism: predictive bias and disparate impact. *Criminology: An Interdisciplinary Journal*, v. 54, n. 4, p. 680-712, 2016. https://psycnet.apa.org/doi/10.1111/1745-9125.12123