UDC 343.98:004.89
DOI https://doi.org/10.24144/2307-3322.2025.91.3.38

THE CODE OF JUSTICE: A HYBRID STANDARD OF PROOF FOR ARTIFICIAL INTELLIGENCE CONCLUSIONS

Shamov O.A.,
Intelligent systems researcher,
head of Human Rights Educational Guild
ORCID: 0009-0009-5001-0526
e-mail: shamov@hreg.org.ua

Shamov O.A. The code of justice: A hybrid standard of proof for artificial intelligence conclusions.

Introduction. The integration of artificial intelligence into justice poses an unprecedented challenge to evidence law. Conclusions generated by opaque algorithms, particularly neural networks, complicate their evaluation using traditional procedural tools. The national doctrine of free evaluation of evidence, relying on the judge's inner conviction, is epistemologically and procedurally unprepared for analyzing such «black boxes», which creates risks of judicial errors and violating the right to a fair trial. Purpose. The article aims to theoretically substantiate and develop the conceptual framework of a hybrid standard of proof for AI-generated conclusions, for its implementation into the procedural legislation of Ukraine. Methods. The research is based on comparative-legal, system-structural, and formal-logical methods. An analysis of the American Daubert/Frye standards and the continental principle of free evaluation of evidence was conducted, allowing for the development of a two-tiered model for evaluating AI evidence. Results. It is established that neither the science-oriented Daubert standard nor the flexible European principle of free evaluation of evidence alone can resolve the «reliability versus transparency» dilemma for AI conclusions. The author's concept of a hybrid standard that harmoniously combines elements of both approaches is proposed. Conclusion. The proposed hybrid model involves a two-tiered procedure: the first stage is a preliminary judicial control («gatekeeping») for scientific validity, reliability, and procedural fairness of the evidence based on adapted Daubert criteria; the second stage is the direct evaluation of the evidence by the judge based on their inner conviction in conjunction with other evidence in the case. This approach allows for the creation of a necessary filter to screen out unreliable algorithmic conclusions while preserving the flexibility of judicial discretion.

Key words: artificial intelligence, evidence law, admissibility of evidence, Daubert standard, free evaluation of evidence, explainable AI (XAI), hybrid model.

Шамов О.А. Код правосуддя: Гібридний стандарт доказування для висновків штучного інтелекту.

Вступ. Інтеграція систем штучного інтелекту в правосуддя створює безпрецедентний виклик для доказового права. Висновки, згенеровані непрозорими алгоритмами, зокрема нейромережами, ускладнюють їх оцінку традиційними процесуальними інструментами. Національна доктрина вільної оцінки доказів, що спирається на внутрішнє переконання судді, є епістемологічно та процедурно непідготовленою до аналізу таких «чорних скриньок», що породжує ризики судових помилок та порушення права на справедливий суд. Метою статті є теоретичне обґрунтування та розробка концептуальних засад гібридного стандарту доказування для висновків, згенерованих ШІ, з метою його імплементації в процесуальне законодавство України. Методи. Дослідження базується на порівняльно-правовому, системно-структурному та формально-логічному методах. Проведено аналіз американських стандартів Daubert/Frye та континентального принципу вільної оцінки доказів, що дозволило розробити дворівневу модель оцінки доказів, отриманих від ШІ. Результати. Встановлено, що ані стандарт Daubert, ані принцип вільної оцінки доказів окремо не здатні ефективно вирішити дилему «надійність проти прозорості» щодо висновків ШІ. Запропоновано авторську концепцію гібридного стандарту, що гармонійно поєднує елементи

обох підходів. Висновки. Запропонована гібридна модель передбачає дворівневу процедуру: перший етап – це попередній судовий контроль («gatekeeping») на предмет наукової валідності, надійності та процедурної справедливості доказу за адаптованими критеріями Daubert; другий етап – це безпосередня оцінка доказу суддею на основі внутрішнього переконання в сукупності з іншими доказами у справі. Такий підхід дозволяє створити необхідний фільтр для відсіювання ненадійних алгоритмічних висновків, водночас зберігаючи гнучкість судового розсуду.

Ключові слова: штучний інтелект, доказове право, допустимість доказів, стандарт Daubert, вільна оцінка доказів, пояснюваний ШІ (XAI), гібридна модель.

Problem Statement. The Fourth Industrial Revolution is relentlessly transforming social relations, and the legal system is no exception. One of the most significant yet least regulated aspects of this transformation is the penetration of artificial intelligence (AI) systems into the realm of evidence. Today, algorithms are used to analyze digital evidence, recognize patterns, identify individuals, assess recidivism risks (e.g., the COMPAS system in the US), and even generate expert conclusions. Whereas computers previously served primarily as tools for storing and processing information, modern AI systems, especially those based on machine learning, can independently form conclusions that can be decisive for a person's fate.

This new reality poses a fundamental challenge to traditional doctrines of evidence law. The Ukrainian procedural system, like most continental law systems, is based on the principle of free evaluation of evidence, enshrined, in particular, in Article 94 of the Criminal Procedure Code of Ukraine, Article 89 of the Civil Procedure Code of Ukraine, and Article 90 of the Commercial Procedure Code of Ukraine. This principle stipulates that no evidence has a predetermined force for the court and is evaluated by the judge's «inner conviction,» based on a comprehensive, complete, and objective examination of all circumstances of the case in their totality.

However, when the evidence is not the testimony of a witness or the opinion of a human expert, but the result of a complex neural network, the concept of «inner conviction» begins to falter. A judge, not being an expert in data science, faces the «black box» problem: they see the input data and the final conclusion, but the logical chain connecting them remains hidden within millions of algorithmic parameters. How can a judge form a reasoned inner conviction about evidence when they can neither understand nor verify the mechanism by which it was obtained? This creates a «procedural gap» [1], undermining the key principles of adversarial proceedings and the right to a defense, as the party against whom such evidence is directed is deprived of a real opportunity to effectively challenge it. Thus, there is an urgent need to adapt national evidence law to the challenges posed by the proliferation of AI and to develop a new standard for evaluating such evidence.

Purpose of the article. The purpose of this article is to theoretically substantiate the necessity and develop the conceptual framework of a hybrid standard of proof for conclusions generated by artificial intelligence, aiming for its further implementation into the procedural legislation of Ukraine. Achieving this purpose involves the following tasks:

- to conduct a comparative-legal analysis of the American Daubert standard for scientific evidence and the continental principle of free evaluation of evidence in the context of their applicability to AI conclusions;
- to identify the strengths and weaknesses of each approach;
- to identify the key works of leading world scholars in this field;
- to formulate the scientific novelty - the concept of a hybrid standard of proof that combines the advantages of both systems;
- to propose practical criteria for assessing the reliability and admissibility of evidence generated by AI in the form of a checklist for judges.

Analysis of Recent Studies and Publications. The problem of the admissibility and evaluation of algorithmic evidence is a subject of active debate in global legal scholarship. A key figure in this field is Professor Andrea Roth, who in her works, particularly «Machine Testimony», points out that algorithms de facto act as «machine witnesses», yet they are mistakenly not subjected to the same reliability checks as human witnesses [2]. She argues that traditional rules of evidence, designed for human perception and error, are inadequate for assessing evidence originating from AI.

Another leading American scholar, John H. Mansfield, a renowned expert on scientific evidence, has dedicated numerous works to the analysis of the Frye and Daubert standards. He emphasizes that

the Daubert standard, although developed for «scientific» expertise, can and should be flexibly applied to all types of technical and specialized knowledge [3], which opens the way for its adaptation to AI conclusions. A key precedent that highlighted the problem was the case of State v. Loomis in Wisconsin (USA), where the court used a recidivism risk assessment generated by the proprietary COMPAS algorithm for sentencing. The state's Supreme Court, while deeming the practice admissible, established a series of caveats, effectively acknowledging that the algorithm's opacity created risks for due process [4]. This case catalyzed a broad academic discussion about the balance between the efficiency of algorithmic tools and individual constitutional rights.

In the European legal field, the discussion centers on the right to explanation and transparency. Researchers, analyzing the GDPR, point out that an effective evaluation of evidence is impossible without understanding the logic of the algorithm's operation [5]. Paul Roberts, a British evidence theorist, emphasizes the fundamental importance of evidential reasoning and how the probabilistic nature of AI conclusions challenges traditional notions of proof «beyond a reasonable doubt» [6]. Singaporean researchers in the article «Artificial Intelligence and Evidence» rightly note that the presumption of reliability traditionally applied to computer records is completely fallacious with respect to complex AI systems, whose conclusions may be biased by the quality of training data [7]. They also draw an analogy between AI conclusions and hearsay evidence, as the algorithm's reliability depends on the data entered and labeled by humans.

Despite a significant number of publications analyzing individual aspects of the problem, a key part of the general problem remains unresolved: the lack of a coherent model that would systematically combine the advantages of the American reliability-oriented approach and the European procedural fairness-oriented approach. Most works either criticize the existing state of affairs or propose piecemeal changes without creating a comprehensive system for evaluating AI evidence for continental legal systems, including Ukraine. This study aims to fill this gap.

Main Results. To develop an effective model for evaluating evidence generated by AI, it is first necessary to analyze in detail the two main approaches that dominate global practice – the American Daubert standard and the European principle of free evaluation of evidence – and then, based on their synthesis, propose a new, hybrid concept.

The American Approach: The Judge as «Gatekeeper» under the Daubert Standard. Historically, the Frye standard (Frye v. United States, 1923) dominated in the US, according to which scientific evidence was admissible if the methodology on which it was based was «generally accepted» in the relevant scientific community. This approach was relatively simple but also conservative, as it blocked access to the court for new but potentially reliable scientific theories.

In 1993, the US Supreme Court in Daubert v. Merrell Dow Pharmaceuticals, Inc. established a new, more flexible and demanding standard. The Court ruled that the judge must act as a «gatekeeper,» actively verifying not only general acceptance but also the scientific validity of the expert opinion before admitting it for consideration by the jury. The Daubert standard, codified in Rule 702 of the Federal Rules of Evidence, offers a non-exhaustive list of criteria for such a check [8]:

Testability: Can the theory or technique be tested (falsified)?

Peer Review and Publication: Has the technique been subjected to peer review and publication?

Known or Potential Rate of Error: What is the known error rate for the given technique?

Existence and Maintenance of Standards: Are there standards controlling the technique's operation?

General Acceptance: Is the technique generally accepted in the relevant scientific community (this Frye criterion was retained, but as one of many, not the sole one).

The advantage of this approach lies in its structured nature and focus on objective indicators of reliability. It forces the party presenting the evidence to prove its scientific validity. However, when applied to AI conclusions, the Daubert standard faces serious obstacles. First, many commercial algorithms (like COMPAS) are «black boxes» due to trade secret protection, making it impossible to verify their internal methodology and error rates. Second, the concepts of «peer review» and «general acceptance» are quite vague for rapidly changing machine learning technologies. As a result, judges are often forced either to reject potentially useful evidence or to accept it on faith, as partially happened in the Loomis case [4].

The European Approach: Flexibility and Risks of «Inner Conviction». The continental system, including the Ukrainian one, is based on the principle of free evaluation of evidence by the judge (or panel of judges). This approach gives the judge considerable flexibility and discretion, allowing them

to consider the totality of the circumstances without formal restrictions on the admissibility of certain types of evidence. Theoretically, the judge must reject evidence if it is not relevant, admissible, credible, and, in aggregate, sufficient. The problem is that with respect to AI conclusions, the judge lacks the tools to assess their credibility. «Inner conviction» must be based on rational analysis, not intuition. If the judge does not understand how the algorithm reached its conclusion, how can they rationally assess its reliability? Moreover, as researchers point out, the opacity of the algorithm violates a party's right to a «voice» - a fundamental element of procedural justice that includes the right to be heard and the right to effectively challenge the opposing party's evidence [1]. When a party cannot cross-examine a «machine witness» or verify its methodology, it is deprived of the opportunity for a full defense. This turns the process into «a trial by an inquisitorial model, where the state has exclusive access to the means of proof» [2].

Therefore, neither approach in its pure form provides an adequate response to the challenges of AI. The American model can be too rigid for proprietary systems, while the European one can be too flexible, creating the risk of legitimizing unreliable and biased conclusions.

Scientific Novelty: A Proposal for a Hybrid Standard of Proof for Ukraine. The solution to this dilemma lies in synthesis. It is proposed to develop and implement into Ukrainian procedural legislation a hybrid standard of proof for conclusions generated by AI, which would combine the American idea of a judicial «gatekeeper» with the flexibility of the continental principle of free evaluation of evidence.

This standard should be a two-tiered procedure:

Tier 1: Preliminary Judicial Gatekeeping for Admissibility. At this stage, the judge, acting as a «gatekeeper» does not evaluate the evidence on its merits but only verifies its compliance with minimum criteria of reliability and procedural fairness. This stage is a mandatory filter. The party submitting the evidence must provide the court with convincing answers to a series of questions, which can be structured as a «judicial checklist» based on adapted Daubert criteria and the principles of Explainable AI (XAI).

- Part A: Reliability and Validity Criteria (adapted Daubert):

Testability and Validation: Has the algorithm undergone independent testing on data relevant to the case at hand? What are the published metrics for precision, recall, and F1-score?

Rate of Error: What is the known rate of false positives and false negatives for the system? Is this rate acceptable for the given category of cases (e.g., requirements in criminal proceedings should be much higher)?

Peer Review and Standardization: Has the underlying architecture of the algorithm and its training methodology been described in peer-reviewed publications? Does the system comply with recognized industry standards (e.g., NIST or ISO standards)?

General Acceptance: Is the application of this type of algorithm (e.g., convolutional neural networks for image recognition) generally accepted in the scientific community for solving similar tasks?

- Part B: Procedural Fairness and Transparency Criteria (based on XAI):

Data Provenance and Bias: On what data was the model trained? What specific measures were taken to identify, audit, and mitigate potential biases (social, gender, racial) in the training dataset? [9]

Explainability («Legal», not technical, explanation): Can the party provide an explanation of the logic for obtaining the specific conclusion that is understandable to a non-specialist (the judge, the other party)? This does not require disclosing the code but requires a justification at the level of «what factors had the greatest impact on the decision» (e.g., using methods like LIME or SHAP) [5]

«Black Box» Audit: If the algorithm is a trade secret, has an independent technical audit report been provided confirming its reliability and impartiality?

Rebutting the Presumption of Reliability: What evidence has been provided to confirm that this AI system is reliable, given that the presumption of reliability for simple computer systems does not automatically extend to it?[7]

If the party cannot provide satisfactory answers to these questions, the judge must find such evidence inadmissible.

Tier 2: Free Evaluation of Evidence in Conjunction with Other Evidence. If the evidence passes the first tier of control and is deemed admissible, it moves to the second tier. Here, the judge applies the traditional principle of free evaluation of evidence: they analyze the AI conclusion not in isolation, but in conjunction with all other evidence in the case (witness testimonies, other expert opinions, physical evidence, etc.) and forms their inner conviction on this basis. At this stage, the judge may give the AI

conclusion more or less weight depending on the circumstances of the case and the reliability of other evidence.

This hybrid model achieves several goals. First, it creates an effective barrier against the use of unreliable, unverified, or biased algorithms in legal proceedings. Second, it protects the procedural rights of the parties by requiring a minimum level of transparency and providing a real opportunity to challenge the evidence. Third, it does not turn the judge into an IT expert but provides them with a structured toolkit to perform their traditional role assessing the reliability of evidence. Thus, the proposed standard does not abolish but strengthens and adapts the principle of free evaluation of evidence to the realities of the digital age [10].

Conclusions. The conducted research has demonstrated that Ukraine's existing model of free evaluation of evidence is insufficient to adequately respond to the challenges associated with the use of conclusions generated by artificial intelligence. Relying on the judge's subjective «inner conviction» without clear objective criteria for assessing the reliability and transparency of such evidence creates significant risks for the justice system. The solution to this problem is the implementation of the proposed hybrid standard of proof. This standard combines the best elements of two leading legal systems: a structured approach to reliability verification, borrowed from the American Daubert standard, and flexibility in the final evaluation of the evidence in the context of the case, which is characteristic of the continental principle of free evaluation. The key element of the proposed model is a two-tiered system:

1. The first tier is a mandatory judicial «gatekeeping» using a checklist that includes criteria for scientific validity, error rates, bias audits, and explainability requirements. This stage acts as a filter that screens out inadmissible «algorithmic evidence.»

2. The second tier is the traditional free evaluation of the evidence that has passed the first tier of control, in conjunction with all other case materials, where the judge determines its evidentiary weight.

This approach allows for a balance between the need to use modern technologies to increase the efficiency of justice and the necessity of protecting fundamental individual rights, particularly the right to an adversarial process and a fair trial. It provides judges with a clear and practical toolkit for working with a new category of evidence, without requiring deep technical knowledge from them, but obliging the party presenting the evidence to prove its reliability and transparency.

Prospects for further research. Further scholarly inquiry in this area should be directed towards developing detailed methodological recommendations for judges and lawyers on the application of the proposed standard to specific types of AI systems (e.g., facial recognition systems, DNA analysis, predictive policing). A separate important direction is the study of the possibility of creating an institution of independent judicial IT auditors who could provide qualified opinions on the reliability and impartiality of algorithms in complex cases. There is also an urgent need to develop and implement corresponding amendments to the procedural codes of Ukraine.

REFERENCES:

1. Kinchin, N. "Voiceless": the procedural gap in algorithmic justice. *International Journal of Law and Information Technology*, *32*(1). 2024. URL: https://doi.org/10.1093/ijlit/eaae024.
2. Roth, A. Machine Testimony. *The Yale Law Journal*. 2017. 126(7). URL: https://ssrn.com/abstract=2893755.
3. Mansfield, J. Scientific Evidence under Daubert. *28 St. Mary's L.J.* 1996. URL: https://commons.stmarytx.edu/thestmaryslawjournal/vol28/iss1/1.
4. Harvard Law Review. State v. Loomis. *Harvard Law Review*. 2017. 130. 1530–1537. URL: https://harvardlawreview.org/print/vol-130/state-v-loomis.
5. Richmond, K., Muddamsetty, S., Gammeltoft-Hansen, T., Olsen, H., & Moeslund, T. Explainable AI and Law: An Evidential Survey. *Digital Society*. 2023. 3(1). URL: https://doi.org/10.1007/s44206-023-00081-z.
6. Roberts, P., & Aitken, K. The Logic of Forensic Proof – Inferential Reasoning in Criminal Evidence and Forensic Science. *ResearchGate.* 2014. URL: https://www.researchgate.net/publication/318596438_The_Logic_of_Forensic_Proof_-_Inferential_Reasoning_in_Criminal_Evidence_and_Forensic_Science.
7. Yeong, W.L. Artificial intelligence and evidence. *Singapore Academy of Law Journal*. 2021. 33. 1–45. URL: https://journalsonline.academypublishing.org.sg/Journals/Singapore-

Academy-of-Law-Journal-Special-Issue/Current-Issue/ctl/eFirstSALPDFJournalView/mid/503/ArticleId/1602/Citation/JournalsOnlinePDF.
8. Cornell Law School Legal Information Institute. (n.d.). Daubert standard. *Wex*. URL: https://www.law.cornell.edu/wex/daubert_standard.
9. Swofford, H., & Champod, C. Implementation of algorithms in pattern & impression evidence: A responsible and practical roadmap. *Forensic Science International: Synergy*. 2021. URL: https://doi.org/10.1016/j.fsisyn.2021.100142.
10. Uriel, D., & Remolina, N. Artificial intelligence at the bench: Legal and ethical challenges of informing – or misinforming – judicial decision-making through generative AI. *Data & Policy*. 2024. 6. https://doi.org/10.1017/dap.2024.53.