JACOPO DELLA TORRE

Decision Rules in the Age of Algorithms

The paper focuses on the relationship between decision rules and algorithms from two complementary perspectives. On the one hand, it analyses how the increasingly broad category of "automatically generated evidence" affects the various conceptions of decision rules developed in legal theory. On the other hand, it explores the use of algorithms as tools assisting both the assessment of evidence and judicial decision-making itself. Particular attention is devoted to the question, long at the centre of a deeply polarized international debate, whether Bayesian methods can contribute to enhancing the rationality, transparency, and accuracy of legal evidentiary reasoning.

Le regole di giudizio nell'era degli algoritmi

L'articolo si concentra sul rapporto tra regole di giudizio e algoritmi da due prospettive complementari. Da un lato, esamina l'impatto della categoria della "prova generata automaticamente" sui diversi significati del concetto di "regola di giudizio". Dall'altro, esplora l'uso degli algoritmi come strumenti di supporto sia alla valutazione della prova sia alla decisione giudiziaria in sé. Particolare attenzione è dedicata alla questione, da tempo al centro di un dibattito internazionale profondamente polarizzato, se i metodi bayesiani possano contribuire a rafforzare la razionalità, la trasparenza e l'accuratezza del ragionamento probatorio penale.

1. *Introduction.* An examination of the relationship between "algorithms" and "decision rules" requires some preliminary terminological clarifications, given the conceptual "indeterminacy"[1] that surrounds both notions.

On the one hand, the term "algorithm" – although it may be broadly defined as a sequence of instructions capable of transforming an input into an output[2] – exhibits a significant degree of functional ambiguity within legal discourse.

[1] On the use of the term "indeterminacy" as a hypernym encompassing different forms of linguistic vagueness, see WALDRON, *Vagueness in Law and Language: Some Philosophical Issues*, in *California Law Review*, 1994, 512.

[2] This explanation is drawn from the document of the *EUROPEAN PARLIAMENTARY RESEARCH SERVICE*, The Impact of the General Data Protection Regulation (GDPR) on Artificial Intelligence, Brussels, 2020, 3, which states that «the concept of an algorithm is more general than the concept of AI, since it includes any sequence of unambiguously defined instructions to execute a task, particularly but not exclusively through mathematical calculations». See also ID., *Artificial Intelligence: How Does It Work, Why Does It Matter, and What Can We Do About It?*, Brussels, 2020, 8, where the notion is defined as «a set of rules defining how to perform a task or solve a problem».

Indeed, as will be discussed below[3], the algorithms employed in criminal proceedings are far from homogeneous[4], differing significantly in their forensic purpose, degree of transparency and verifiability, and the level of decisional autonomy entrusted to them.

Nor is greater conceptual clarity achieved by limiting the analysis to algorithms falling within the notion of "artificial intelligence" (hereinafter, "AI")[5]. This is because even the principal definitions of "AI" currently provided at the normative level – such as that contained in Article 3(1), no. 1, of Regulation (EU) 2024/1689 (the so-called "AI Act")[6], incorporated into Italian law by virtue of the reference contained in Article 2(1)(a) of Law No. 132 of 23 September 2025[7] – are drafted in terms so expansive that they provide only limited guidance[8].

---

[3] See *infra* § 2.

[4] In this regard, see. QUATTROCOLO, *Prove e intelligenza artificiale*, in *La prova scientifica*, Conti-Marandola (eds.), Milan, 2023, 477 ff.

[5] As is well known, the concept of "artificial intelligence" is polysemous. For example, a distinction is commonly drawn between a theoretical model of a "strong AI", which is understood as a form of artificial intelligence endowed with general cognitive capacities comparable to those of human beings, and a "weak AI", which instead refers to the systems currently in existence, designed to perform specific tasks through techniques for processing input data. However, even within the latter genus it is possible to find highly heterogeneous applications, which may – through successive approximations – be classified according to different levels of complexity: these range from programs based on predefined rules (rule-based systems), in which the outcome constitutes the deterministic result of explicit instructions predetermined by the programmer, to systems grounded in machine-learning techniques, capable of identifying correlations within data and progressively adapting their own functioning. For a technical analysis of the issues raised by this notion, see FLORIDI, *Etica dell'intelligenza artificiale. Sviluppi, opportunità, sfide*, Milan, 2022, 40-41.

[6] Pursuant to this provision, an "AI system" shall be understood as «a machine-based system designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that may influence physical or virtual environments». As is well known, the issue of defining the concept of AI has been one of the most contentious matters throughout the preparatory works of the AI Act. For a summary in this regard, see the document of *EUROPEAN LAW INSTITUTE, Commission Guidelines on the Application of the Definition of an AI System and the Prohibited AI Practices Established in the AI Act. Response of the European Law Institute*, Wien, 2024, 9 ff., which also provides a comparative analysis with the definitions contained in other legislative instruments.

[7] Containing "provisions and delegations to the Government concerning artificial intelligence". As regards the interplay between this law and the AI Regulation, see, from a criminal procedural perspective, CANZIO, *Investigations and Trial between Regulation (EU) 2024/1689 (AI Act) and Law No. 132 of 2025*, in www.giustiziainsieme.it, 13 January 2026.

[8] The definition contained in the AI Act has been described as «very broad, if not generic» by CARTA, *Il Regolamento UE sull'Intelligenza Artificiale: alcune questioni aperte*, in *Eurojus*, 2024,

On the other hand, in the absence of a precise legislative definition, even the notion of "decision rule" is inherently ambiguous, as it may be understood in at least three distinct ways[9].

According to an authoritative interpretation, the term may, broadly speaking, encompass all evidentiary rules relevant to fact-finding, including those governing the admission, exclusion, and evaluation of evidence[10]. In an intermediate sense, it may refer more specifically to the rules governing the assessment of evidence, whether directed at particular categories of proof[11] – for example, the rules set out in Articles 192(2) and (3) of the Italian Code of Criminal Procedure (CCP) – or the evidentiary framework as a whole[12]. Finally, in its narrowest sense, the notion denotes the rule that determines the outcome of the case when the available evidence leaves the relevant facts uncertain, thereby allocating the risk of non-proof to one of the subjects involved in the proceedings[13].

Moreover, even when the analysis is confined to the narrowest meaning of the expression "decision rule", the conceptual picture remains only partially clear. In fact, a central component of any decision rule is the standard of proof[14], namely the rule that determines the minimum degree of evidentiary support

---

3, 200, fn. 56. The same conclusion applies, moreover, to the substantially similar definition contained in Article 2 of the Council of Europe Framework Convention on Artificial Intelligence (CETS 225 – Artificial Intelligence, 5.IX.2024).

[9] The following discussion proceeds on the basis of the approach proposed by FERRUA, *Regole di giudizio (Dir. pen. proc.)*, in *Enc. dir.*, *Annali* X, Milan, 2017, 726. For deeper insights on the subject, see DINACCI, *Regole di giudizio* (*Dir. proc. pen.*), in *Dig. disc. pen.*, Suppl. VIII, Turin, 2014, 644 ff.; PAULESU, *La presunzione di non colpevolezza dell'imputato*, Turin, 2009, 179 ss. and ORLANDI, *Discrezionalità e regole di giudizio nel processo penale*, in *Dir. pen. cont. – Riv. trim.*, 2024, 4, 21.

[10] This is, for example, the first hypothesis formulated by FERRUA, "*Regole di giudizio*", cit., 726.

[11] As regards this *genus*, see, for all, DANIELE, *Regole di esclusione e regole di valutazione della prova*, Turin, 2009.

[12] In this sense, see, for example, MAZZA, *Il ragionevole dubbio nella teoria della decisione*, in *Criminalia*, 2012, 363.

[13] See, in this regard, FERRUA, *Brevi appunti in tema di udienza preliminare, appello e improcedibilità*, in *Discrimen*, 9 December 2021, 2, and PAULESU, "*Presunzione di non colpevolezza*", in *Dig. disc. pen.*, vol. IX, Turin, 1995, 684.

[14] On the connection between the category of standard of proof and decision rules, cf. TUZET, *Assessment criteria or standards of proof? An effort in clarification*, in *Artificial Intelligence and Law*, 2020, 96 f. and DELLA TORRE, *Gli* standard *di prova per la condanna penale tra storia e attualità*, in www.lalegislazionepenale.eu, 14 November 2023, 6. As regards, instead, a comparison with the concept of burden of proof, see ALLEN, *Burdens of Proof*, in *Law, Probability and Risk*, 2014, 13(3-4), 195 ff., as well as STEIN, *Foundations of Evidence Law*, Oxford, 2005, 118 ff.

required for a factual allegation to be regarded as proven[15]. In this respect, standards of proof operate within the protasis of decision rules, insofar as they define the conditions triggering their application. More specifically, they establish the level of evidentiary corroboration necessary for a factual proposition to be treated as proven and, conversely, identify the circumstances in which the proposition must still be regarded as affected by residual evidentiary uncertainty.

According to the prevailing view, however, standards of proof in criminal proceedings cannot be reduced to a single evidentiary threshold. Rather, they are articulated into a plurality of differentiated standards governing distinct procedural decisions[16]. Firstly, there are "incidental standards", which are used to determine whether a given factual assertion can be accepted as proven within a specific sub-phase of the proceedings[17]. Secondly, there are "propulsive standards", which determine the minimum evidentiary basis required to initiate or advance criminal proceedings[18]. Finally, there are "strictly decisional standards", which guide the court in determining whether the evidence is sufficient to justify a conviction[19].

---

[15] The literature on standards of proof is extensive. For an overview of the principal doctrinal debates on the subject, see DELLA TORRE, *Taking the Evolution of the Standard of Proof for a Criminal Conviction Seriously*, in *Quaestio Facti*, 2025, 156 as well as the recent monograph by FERRER BELTRÁN, *Prova senza convincimento*. Standard *di prova e giusto processo*, Turin, 2025.

[16] On this point, see, in Italian literature, CAPRIOLI, "*Condanna*" *(dir. pen. proc.)*, in *Enc. Dir., Annali* II, vol. I, Milan, 2008, 112; GAROFALO, *La diversificazione degli* standard *di prova nel processo penale e nel rapporto fra giurisdizioni*, in *Cass. pen.*, 2020, 3891 ff.; ORLANDI, *Discrezionalità e regole di giudizio*, cit., 20 ff.; TUZET, *La prova ragionata*, Milan, 2023, 226 ff.; VIGONI, *Giudizi prognostici e ragionevole dubbio*, in *Giudizio penale e ragionevole dubbio*, ed. by Incampo-Scalfati, Bari, 2017, 378 ff. For an authoritative opposing view, however, see FERRUA, *Onere della prova, regola di giudizio* e standard *probatorio: alla ricerca della perduta proposizione da provare*, in *Cass. pen.*, 2020, no. 7, 2639 ff., according to whom the only standard of proof would be that of "beyond a reasonable doubt", since below such threshold the very concept of proof would cease to make sense.

[17] The paradigmatic example is the standard of "serious indications of guilt" required for the imposition of a precautionary measure, as provided for by Article 273 CCP. On this point, see, also for further bibliographical references, VIGONI, *Giudizi prognostici e ragionevole dubbio*, cit., 390.

[18] For instance, one may consider the standard of the "reasonable prospect of conviction" (for a deeper analysis, see CONTI, *La ragionevole previsione di condanna: verso un sistema accusatorio contemporaneo*, in *Dir. pen. proc.*, 2024, no. 12, 1543, also for the relevant doctrinal literature cited therein), codified by Legislative Decree No. 150 of 10 October 2022 in order to govern the choice between prosecution and dismissal (Article 408 CCP), as well as between the starting of trial and a judgment of no ground to proceed (Articles 425 and 554-*ter* CCP).

[19] In this regard, from a criminal law perspective, the reference goes to standards such as "beyond a reasonable doubt", codified in Article 533 CCP, as well as to those of "moral certainty" and *intime convic-*

The foregoing classification makes it particularly apparent that the potential interaction between AI algorithms and decision rules may concern a wide range of legal concepts and procedural mechanisms. Their common denominator lies, for the most part, in the influence that computational tools may exert on the law of evidence and, more broadly, on judicial fact-finding. Yet such a broad conception ultimately offers limited analytical value.

For this reason, the present study deliberately confines its focus to two specific perspectives, which will structure the analysis developed in the following pages.

On the one hand, it examines how the increasingly broad category of "automatically generated evidence"[20] produced through AI algorithms affects the various, broader or narrower, conceptions of the term "decision rule" outlined above[21].

On the other hand, particular attention is devoted to the use of algorithms as tools assisting both the assessment of evidence, whether AI-generated or not, and judicial decision-making itself. From this perspective, the analysis focuses especially on the question – long at the centre of a deeply polarized international debate[22] – whether Bayesian methods can contribute to improving the rationality, transparency, and accuracy of legal evidentiary reasoning[23].

What unites these two perspectives is the recognition that decision rules within the European legal landscape are already being reshaped by the emergence of AI tools. Although current judicial practice provides no support for genuine forms of "robot judges", as confirmed by the CEPEJ[24], algorithmic tools are already being employed in tasks that increasingly approach the core of legal reasoning. Precisely for this reason, it becomes imperative to devise ade-

---

tion (for a deeper analysis see – also for further bibliographical references – DELLA TORRE, *Taking the Evolution*, cit., 156 ff.). From a civil law perspective, by contrast, relevant standards are "preponderance of the evidence" and the presence of "clear and convincing evidence". For a distinction between criminal and civil standards of proof, see the synthesis provided by TUZET, *La prova ragionata*, cit., 230 ff.

[20] On this point, cf. QUATTROCOLO, *Prove e intelligenza artificiale*, cit., 477 ff.

[21] On this point, see *infra*, §§ 2 – 2.3.

[22] Which is effectively resumed by LUPÁRIA DONATI, Trials by probabilities. *Qualche annotazione "eretica"*, in *L'uso della prova scientifica nel processo penale*, ed. by Gucci-Gennari-Gentilomo, Santarcangelo di Romagna, 2012, 95 ff.

[23] In this regard, see §§ 3 – 3.2.

[24] On this point, see the document published by CEPEJ-AIAB (2024) 4 Rev5, Strasburg, 28 February 2025, 12, where, at the end of an articulated empirical research, it is explicitly stated that, «according to the available information, there are no fully automated AI systems that could function entirely independently in the courts (so called "robot-justice")».

quate safeguards capable of ensuring that computational technologies fulfil their potential to enhance the administration of justice, rather than becoming a source of algorithmic distortions liable to compromise the guarantees of a fair trial[25]. And it is equally apparent that this challenge is particularly complex in the Italian criminal justice system, which has long been marked by difficulties in developing an adequate regulatory framework for the use of new technologies in criminal proceedings[26].

2. *AI generated evidence and decision rules.* The fact that, in recent years, AI has become one of the cornerstones of contemporary societies is evident to everyone. «Education, business and industry, travel and logistics, banking, retail and shopping, entertainment, welfare and healthcare, politics and social relations: in short, life itself as we know it today has become inconceivable»[27] without the presence of «artificial agents»[28]. The outcomes of this transformative process are striking even from a qualitative perspective, as AI systems have begun to match – and, in some cases, surpass – human performance in highly technical and creative tasks, including those requiring sophisticated clinical reasoning[29].
Naturally, the field of justice in general, and criminal justice in particular, could not remain unaffected by this phenomenon[30]. The influence of AI on this domain is so pervasive that it can be effectively illustrated through circular

[25] MAZZA, *Distopia del processo artificiale*, in *Arch. pen. web*, 2025, 1, 1 considers this risk particularly concrete.
[26] In this regard, cf. DELLA TORRE, *Cripto-attività e giustizia penale*, Milan, 2026, 410 and the bibliography cited therein.
[27] The quotation is taken from FLORIDI, *Etica dell'intelligenza artificiale*, cit., 11.
[28] The term is taken from FLORIDI-SANDERS, *On the Morality of Artificial Agents*, in *Minds and Machines*, 2004, 14(3), 349.
[29] Cf., for instance, the recent study published by BRODEUR (and others), *Performance of a large language model on the reasoning tasks of a physician*, in *Science*, 30 April 2026, 524 ff.
[30] For a resume of the vast debate on this matter, cf.: BACCARI-FELICIONI (eds.), *La decisione penale tra intelligenza emotiva e intelligenza artificiale*, Milan, 2023; BELVINI, *Intelligenza artificiale e circuito investigativo*, Bari, 2025; DAMOSSO, *Il giudice profilabile. Studio sull'impatto dei* Judge Profiling A.I. Tools *nel processo penale*, Turin, 2025; FIORELLI, *Calcolabilità del giudizio giuridico e processo penale*, Turin, 2025; LELIEUR (ed.), *Artificial Intelligence and Administration of Criminal Justice*, special number to the *Revue Internationale de Droit Pénal*, 2023; NIEVA FENOLL, *Intelligenza artificiale e processo*, Turin, 2019; QUATTROCOLO, *Artificial Intelligence, Computational Modelling and Criminal Proceedings. A Framework for A European Legal Discussion*, Cham, 2020.

diagrams showing how these technologies are deployed throughout the entire criminal process, from the pre-trial phase to the execution of the sentence[31].
It is widely acknowledged that, alongside predictive risk-assessment algorithms (the so-called "risk assessment tools")[32], one of the areas in which the impact of digital technologies and AI tools has manifested itself most intensely is that of the so-called "AI-based evidence"[33]. By this expression, we refer to the set of evidentiary elements generated by computational systems through the automated processing of input data according to specific algorithmic models[34]. Evidently, this is a highly heterogeneous category, encompassing both tools specifically designed for investigative or forensic purposes[35] and the procedural use of information derived from applications originally developed for general-purpose functions[36].
The emergence of this category of evidence is hardly surprising, as it is closely connected to the phenomenon commonly described as the "*data deluge*" that characterizes contemporary societies[37]. Indeed, digital technologies make it possible to select and analyze the enormous mass of data generated by information systems, transforming it into an "open-pit mine"[38] of elements that can be examined in extremely short timeframes and with such a level of granulari-

[31] Cf. EPIC, *Ai in the Criminal Justice System*, in *www.epic.org*. For a representation of the different uses of AI, from an Italian perspectice, see SCALFATI, *IA e processo penale: prospettive d'impiego e livello di rischio*, in *Proc. pen. giust.*, 2024, 1405 ff.
[32] The debate on this point, which developed also following the global mediatic echo of the judgement delivered by the Supreme Court of the State of Wisconsin (Wisconsin *State v. Loomis*, 881 NW 2d 749 (WIS 2016)), is undoubtedly vast. For an Italian resume of such debate, see GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei* risk assessment tools *tra Stati Uniti ed Europa*, in *Dir. pen. cont.*, 29 May 2019; MALDONATO, Risk and need assessment tools *e riforma del sistema sanzionatorio: strategia collaborative e nuove prospettive*, in *Intelligenza artificiale e processo penale*, ed. by Di Paolo-Pressacco, Naples, 2022, 146 ff.; QUATTROCOLO, *Sui rapporti tra pena, prevenzione del reato e prova nell'era dei modelli computazionali psico-criminologici*, in *Teoria e critica della regolazione sociale*, 2021, 1, 264 ff.
[33] On the relationship between algorithms and criminal evidence, see, in Italian literature, the pioneristic work by LUPÁRIA DONATI, *Prova giudiziaria e ragionamento artificiale: alcune possibili chiavi di lettura*, in *Il concetto di prova alla luce dell'intelligenza artificiale*, ed. by Sallantin-Szczeciniarz, Milan, 2005, XIV ff.
[34] The aspect of the automated processing of available data in defining the concept of "evidence" generated automatically is emphasized by QUATTROCOLO, Prove e intelligenza artificiale, cit., 479.
[35] For a heterogeneous range of examples, see CAMON, *Intelligenza artificiale, indagini preliminari e diritti di libertà*, in *Cass. Pen.*, 2025, 3006 ff.
[36] On this point, see again QUATTROCOLO, *Prove e intelligenza artificiale*, cit., 478 f.
[37] See the title given to the special number of THE ECONOMIST, *The data deluge*, 25 February 2010.
[38] The term is taken from RODOTÀ, *Trasformazioni del corpo*, in *Pol. dir.*, 2006, 1, 6.

ty as to allow the identification of patterns that would otherwise be difficult for the human operator to detect. Consequently, such a socio-cultural environment –one in which individuals are continuously encouraged to generate data within an uninterrupted informational circuit – contributes to granting algorithmic evidence ever-increasing significance, to the extent that, in certain fields, it may come to assume the role of the contemporary era's new "queen of evidence"[39].

Nevertheless, despite its considerable potential, the risks associated with this phenomenon must not be overlooked. Such concerns have been already highlighted by extensive scholarly literature[40], as well as by the first soft law[41] and hard law[42] instruments adopted on this matter.

In this respect, it is worth noting that AI evidence, beyond its potential to affect a wide range of traditional fundamental rights and freedoms[43], raises several specific challenges[44].

Among these concerns, particular attention must be paid to the heightened risk of generating the so-called "automation bias"[45], namely the tendency to at-

[39] This is the case, for instance, with software outputs generated for blockchain analysis in criminal proceedings involving the tracing of crypto-asset flows; on this point, cf. DELLA TORRE, *Cripto-attività e giustizia penale*, cit., 249 ff.

[40] Cf. the pioneering monographic study by QUATTROCOLO, *Artificial Intelligence*, cit., *passim*.

[41] Here the reference goes, first of all, to the CEPEJ Ethical Charter, on the use of IA in judicial systems and in other connected areas, CEPEJ (2018) 14, with the comment by QUATTROCOLO, *Intelligenza artificiale e giustizia: nella cornice della Carta etica europea, gli spunti per un'urgente discussione tra scienze penali e informatiche*, in *www.lalegislazionepenale.eu*, 18 December 2018. See also the EU Parliament Resolution of 6 October 2021, on IA in criminal law and its use by police and judicial authorities, P9_TA (2021) 0405. For a deeper analysis, see BARONE, *Intelligenza artificiale e processo penale: la linea dura del Parlamento europeo*, in *Cass. pen.*, 2022, 118 ff.

[42] Besides the already cited AI Act (for a criminal procedure-oriented comment on the matter, see QUATTROCOLO, *Intelligenza artificiale e processo penale: le novità dell'*AI Act, in *Dir. dif.*, 16 gennaio 2025), the reference obviously goes in particular to the European Framework Convention in IA, CETS no. 225 of 5 November 2024.

[43] As it is established by Recital 48 of the AI Act, such rights include at least «the right to human dignity, respect for private and family life, the protection of personal data, freedom of expression and information, freedom of assembly and association, and the right to non-discrimination, the right to education, consumer protection, workers' rights, the rights of persons with disabilities, gender equality, intellectual property rights, the right to an effective remedy and to a fair trial, the rights of the defence and the presumption of innocence, and the right to good administration».

[44] And this is in line with the general idea according to which the IA phenomenon «exacerbates old ethical problems, reshapes some of them, and creates new ones» (cf. FLORIDI, *The Ethics of Artificial Intelligence: exacerbated problems, renewed problems, unprecedented problems*, special number of *American Philosophical Quarterly*, 2024, 301).

tribute to the outputs of algorithmic systems a degree of reliability and objectivity that is not adequately supported by empirical evidence[46]. Such a phenomenon may encourage excessive reliance on algorithmic outputs, leading decision-makers to assign them a weight disproportionate to the role they should properly play within the adjudicative process[47]. It must, of course, be acknowledged that similar problems also characterize the use of traditional forms of scientific evidence. In the context of algorithmic evidence, however, this risk appears particularly acute, owing both to the automated nature of the systems involved and to the widespread perception that computational processes are inherently neutral and objective.

Moreover, attention must also be paid to the so-called "black box problem", namely the risk that algorithmic systems may operate through decision-making processes characterized by a significant degree of opacity[48]. As is widely acknowledged, it is often difficult – and in some cases virtually impossible – to reconstruct the logic underlying certain algorithms[49] and, consequently, to understand how particular inputs are transformed into specific outputs[50]. This lack of transparency raises serious concerns, from an evidentiary perspective, regarding both the intersubjective reviewability of algorithmic outputs and the rational justification of judicial decisions[51].

Finally, perhaps even more insidious is the problem of systemic bias[52], namely structural distortions stemming from training data and/or from the design choices embedded in computational systems, which may produce particularly

---

[45] In this regard, see PERRONE, *La prognosi postuma tra distorsioni cognitive e* software *predittivi*, Turin, 2021, 117.

[46] On this point, see LUPÁRIA DONATI-FIORELLI, *Diritto probatorio e giudizi criminali ai tempi dell'intelligenza artificiale*, in *Dir. pen. cont., Riv. trim.*, 2022, 2, 40.

[47] On the role that the proportionality principle should instead perform within this matter, see PULITO, *Il principio di proporzionalità algoritmica*, in *Giudizio penale e intelligenza artificiale*, ed. by di Incampo-Triggiani, Bari, 2026, 245 ff.

[48] On this point, see ADENSAMER-KLAUSNER, "*Part Man, Part Machine, All Cop": Automation in Policing,* in *Front. Art. Intell.*, 23 June 2021, 4.

[49] This category includes many applications of machine learning/deep learning algorithms, which therefore constitute the principal source of concern: see LISI, *Intelligenza Artificiale Affidabile: dalle applicazioni in ambito forense alla simbiosi fra umani e macchine intelligenti*, in *Giudizio penale e intelligenza artificiale*, ed. by Incampo–Triggiani, cit., 88).

[50] In this regard, see CONTISSA-LASAGNI-SARTOR, *Quando a decidere in materia penale sono (anche) algoritmi e IA: alla ricerca di un rimedio effettivo*, in *Dir. internet*, 2019, 4, 620.

[51] See, once again, LUPÁRIA DONATI-FIORELLI, *Diritto probatorio e giudizi criminali*, cit., 41.

[52] In this regard, see VALENTINI, *La "scoperta" dei* bias *cognitivi nel processo penale*, in *Arch. pen. web*, 2025, 3, 2 ff.

serious effects, at times even of a discriminatory nature[53]. This problem is further exacerbated by the fact that algorithms, by their very nature, tend to replicate and amplify the patterns contained in the data on which they are trained, thereby giving rise to phenomena of "error recursivity"[54], whereby original distortions are not merely reproduced but progressively consolidated and reinforced, in accordance with the well-known "garbage in, garbage out" logic[55].
These brief remarks are sufficient to explain why the legal community advocates a particularly cautious approach to the use of AI in the criminal justice domain, to ensure that such tools do not ultimately undermine fundamental rights and freedoms and, in particular, fair trial guarantees[56]. Such an approach requires not only reflection on how the safeguards traditionally developed for technical and scientific evidence may be extended to algorithmic evidence, but also the introduction of specific protections aimed at addressing the distinctive risks posed by these technologies. Among these safeguards, particular importance must be attached to preserving a relationship of cooperation, rather than substitution, between computational tools and human actors, with the latter necessarily retaining ultimate decisional control and responsibility[57].
It is precisely from this perspective that the following sub-sections should be read. Their purpose is to demonstrate how decision rules – in their progressively narrower meanings – are, if taken seriously, capable of providing three concentric layers of "human safeguards" of considerable importance in addressing the risks associated with the use of algorithms in the criminal justice domain.

---

[53] Cf., for instance, NAJIBI, *Racial Discrimination in Face Recognition Technology*, in *www.sitn.hms.harvard.edu*, 24 October 2020; WAELEN, *The struggle for recognition in the age of facial recognition technology*, in *AI and Ethics*, 8 March 2022, 4.
[54] In this regard, see DELLA TORRE, *Cripto-attività e giustizia penale*, cit., 364.
[55] In this regard, see EBSCO, *Garbage in, garbage out (GIGO)*, in *www.ebsco.com*.
[56] Cf., for instance, the approach advanced by LUPÁRIA DONATI–FIORELLI, *Diritto probatorio e giudizi criminali*, cit., 42 ff.
[57] See, in particular, CANZIO, Ai Act *e processo penale: sfide e opportunità*, in *Giudizio penale e intelligenza artificiale*, ed. by Incampo–Triggiani, *op. cit.*, p. 141, who «recognises as admissible within the sphere of criminal jurisdiction only the 'weak' and 'collaborative' model of artificial intelligence which, within the framework of human–machine complementarity, nonetheless allows human beings to retain control over the machine».

2.1. *The judicial review on the admissibility of AI evidence.* The admissibility of AI-based evidence is among the issues that have attracted the greatest attention in the debate surrounding the use of computational tools in the administration of justice[58]. This is unsurprising, given that the establishment of robust "entry barriers" represents the most effective safeguard against the risk of a technocratic drift resulting from excessive reliance on computational systems[59]. Not coincidentally, this issue is also of central importance from the perspective adopted in this article, insofar as the relevant rules contribute to defining the range of evidentiary elements upon which the decision-maker may legitimately rely and, consequently, the sphere within which the narrower meanings of the concept of decision rule operate.

For the present work, particular attention should focus on two tasks that the judge is required to perform in this respect.

The first concerns the lawfulness of the AI application itself. Pursuant to the general rules of the Italian legal system, the judge must ascertain whether the specific tool whose output is intended to be used is prohibited by law, having regard both to domestic[60] and to supranational sources and, in particular, whether – due to the unacceptable level of risk it poses to fundamental rights – it falls within the category of practices prohibited under Chapter II, Article 5 of Regulation (EU) 2024/1689[61]. From this perspective, it should be noted that, although the AI Act is primarily concerned with market regulation[62] and

[58] See, for instance, *ex multis*, CANZIO, *Intelligenza artificiale, algoritmi e processo penale*, in *www.sistemapenale.it*, 8 gennaio 2021; ID., *Intelligenza artificiale e processo penale*, in *Prova scientifica e processo penale*, ed. by Canzio-Lupária Donati, 2ª ed., Milan, 2022, 910 ff.; QUATTROCOLO, *Prova e intelligenza artificiale*, cit., 484 ff.; UBERTIS, *Perizia, prova scientifica e intelligenza artificiale nel processo penale*, in *www.sistemapenale.it*, 3 June 2024, 16 f.

[59] In this sense, cf. LUPÁRIA DONATI-FIORELLI, *Diritto probatorio e giudizi criminali*, cit., 42 f.

[60] The reference goes, for instance, to AI tools that are used to produce evidence in contrast with the prohibitions set out in Arts. 188, 64, par. 2 or 220, par. 2, CCP.

[61] As is well known, Article 5 of the AI Act encompasses several categories of particular relevance to criminal procedure, including, in particular: (1) the placing on the market or use of AI systems intended to carry out risk assessments relating to natural persons in order to predict the risk of their committing a criminal offence, solely on the basis of the profiling of a natural person or the assessment of personality traits and characteristics (Art. 5(1)(d)); (2) the placing on the market, putting into service for that specific purpose, or use of AI systems that create or expand facial recognition databases through the untargeted scraping of facial images from the internet or CCTV footage (Art. 5(1)(e)); and (3) the use of real-time remote biometric identification systems in publicly accessible spaces, unless such use is strictly necessary to pursue specific interests deemed by the European legislature to warrant enhanced protection (Art. 5(1)(h)).

[62] This point is emphasized by QUATTROCOLO, *Intelligenza artificiale e processo penale*, cit.

the European institutions have traditionally displayed considerable reluctance to introduce strict exclusionary rules of evidence[63], this regulatory framework may nevertheless give rise to significant preclusive effects within criminal proceedings[64], particularly when read in conjunction with Article 191 of the Italian CCP. Three arguments support this conclusion. The first is textual in nature, since Article 5 of the AI Act establishes genuine prohibitions on the use of certain AI practices, which may therefore fall within the notion of "evidentiary prohibitions established by law" for the purposes of Article 191 of the CCP. The second is teleological, insofar as the very purpose of the provision is to preclude the admissibility of particularly harmful practices. The third argument derives from the principle of primacy of EU law: denying such consequences from the standpoint of evidentiary inadmissibility would risk depriving the protection afforded by the European legislature of its practical effectiveness (*effet utile*), thereby undermining the objectives pursued by the Regulation[65].

This issue, however, must be distinguished from a second and conceptually separate inquiry concerning the admissibility of algorithmic output as evidence. At least pending the implementation of the domestic delegations concerning the AI Act for investigative purposes under Law No. 132/2025[66], the decision-maker must still assess, on a case-by-case basis, whether the admis-

[63] In this regard, see, for all, CABIALE, *I limiti alla prova nella procedura penale europea*, Milan, 2019, *passim*.

[64] As a matter of fact, the AI Regulation itself, in certain instances – for example, with regard to real-time biometric identification tools (Art. 5(3)) – provides for the obligation to immediately cease the use of the outputs generated by computational systems in the absence of specific legitimizing conditions.

[65] About the importance to always ensure the respect of the principle of effectiveness of European Union law, see, from a criminal procedure and evidence-based perspective, Court of Justice of the EU, *La Quadrature du Net e a.,* 6 October 2020, C-511/18, C-512/18 e C-520/18, parr. 223-227.

[66] The reference goes to the fact that this legislative act contains two particularly significant enabling provisions, namely, on the one hand, Article 24(2)(h), concerning the regulation of the use of AI systems by law enforcement authorities; and, on the other hand, Article 24(5)(e), which provides for the regulation of the use of AI tools in the investigative phase, «in compliance with the safeguards inherent in the right of defence and personal data protection, as well as the principles of proportionality, non-discrimination and transparency». On this point, see DELLA TORRE, *Cripto-attività e giustizia penale*, cit., 466 ff. It should also be noted that, on 10 June 2026, the Italian Government approved, on a preliminary basis, a draft legislative decree establishing harmonised rules on artificial intelligence, including provisions on the use of AI systems for law-enforcement purposes and related issues of civil and criminal liability. The draft appears to regulate only certain uses of AI in policing and criminal investigations. Consequently, for situations falling outside the scope of those specific provisions, Article 189 of the CCP is likely to retain a residual yet crucial role in governing the admissibility of algorithmic evidentiary tools.

sion of the output of an algorithmic tool falls within the scope of Article 189 of the CCP[67]. As is well known, this provision requires, following an anticipatory adversarial debate, a twofold assessment. On the one hand, the judge must evaluate the epistemic suitability of the evidentiary means, namely its capacity to provide «a result that is concretely usable (also because it is rationally controllable and justifiable) for the verification of the individual proposition forming the subject matter of proof»[68]. On the other hand, the judge must ascertain its compatibility with the moral freedom of the person involved in its acquisition[69].

That being said, it should be noted that, for the purposes of carrying out such scrutiny, a decisive role is played by the application – if necessary, also through the appointment of one or more experts so as to render the adversarial proceeding effective[70] – of two broad decision rules specifically designed to contain the risks arising from AI applications.

First, reference must be made to compliance with an adequate standard of "algorithmic transparency"[71], which is widely recognized as one of the fundamental pillars both of the European regulatory framework on AI[72] and, at the domestic level, of Law No. 132/2025[73]. For present purposes, this requirement must be construed as meaning that sufficient information must be available to enable the judicial authority to assess, already at the admissibility stage of the evidence, whether an adequate level of comprehensibility and explainability of the output generated by the system has been ensured. In this regard, it should be clarified that such a safeguard cannot be reduced to a merely technical requirement but rather takes the form of a structural precondition

[67] See, in particular, CANZIO, *Intelligenza artificiale e processo penale*, cit., 910 and QUATTROCOLO, *Prova e intelligenza artificiale*, cit., 492 f.

[68] See UBERTIS, *Perizia, prova scientifica e intelligenza artificiale*, cit., 11.

[69] For a recent analysis of this criterion, see, CAMON, *Processo e libertà morale: tendenze recenti della giurisprudenza, della dottrina, della legislazione*, in *Proc. pen. giust.*, 2024, 4, 1026 ff., as well as the study by CHELO, *La tutela della libertà morale nella formazione della prova penale,* Milan, 2025.

[70] The idea is drawn from DELLA TORRE, *Cripto-attività e giustizia penale*, cit., 457 ff.

[71] In this regard, cf. QUATTROCOLO, *Prova e intelligenza artificiale*, cit., 481 f.

[72] The principle of transparency was first recognized in the fourth criterion set out in the CEPEJ Ethical Charter on the use of artificial intelligence in judicial systems and is now codified in Article 13 of the AI Act.

[73] As a matter of fact, the principle of transparency appears throughout this statute, starting from Article 1, concerning "purpose and scope", and Article 3, which sets out a series of "general principles".

of equality of arms and adversarial proceedings[74]. Indeed, the issue extends beyond merely granting access to the source code, which is already rendered particularly problematic by the widespread use, in many fields, of software protected by trade secrets[75]. Instead, it must be understood in a substantive sense, encompassing the requirements of explainable AI, namely the preference for systems capable of rendering intelligible to human understanding the logical pathway leading from the input data to the final output[76].

Nevertheless, particularly in a field as sensitive as criminal justice, ensuring the full effectiveness of the transparency principle requires going further by demanding that users be placed in a position to know, already at the admissibility stage, the expected level of accuracy of the system and the factors capable of affecting its performance. As recognized by Article 13 of the AI Act with regard to high-risk systems, this entails, in practical terms, an obligation to disclose the conditions under which the algorithmic output may prove less reliable, the types of data likely to generate distortions or false positives, as well as the variables affecting the probability of error, all of which must be capable of being known and assessed. In this respect, it is important to stress the affinity between the safeguards outlined above and the so-called *Daubert* criteria governing the validation of scientific and technical evidence[77], which must be taken seriously in this context as well[78], as a safeguard against forms of algorithmic pseudo-science[79].

---

[74] In this regard, cf. QUATTROCOLO, *Artificial Intelligence*, cit., 93 ff., as well as ERDOGAN, *Diving the Iceberg: Establishing Transparency in AI for Law Enforcement*, in *European Papers*, 2024, 9(3), 956 ff.

[75] For a concrete example about the problems created by this *status quo*, see the US case *USA v. Sterlingov*, 2024 WL 860983, *Criminal Action* No. 21-399 (RDM) (D.D.C. Feb. 29, 2024). As is well known, a commendable line of case law developed by the Italian Council of State (*Consiglio di Stato*) (see Cons. Stato, sez. VI, 4 February 2020, n. 881, in *One Legale*) has affirmed the principle that software developers cannot invoke confidentiality concerns in order to deny individuals access to the source code of an algorithm used by a public authority.

[76] In this regard, see SAMEK (and others), *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*, Cham, 2019, *passim*; RAI, *Explainable AI: from black box to glass box*, in *Journal of the Academy of Marketing Science*, 2020, 48, 137 ff.; ZORNOZA, *Explainable Artificial Intelligence*, in *www.towardsdatascience.com*, 15 April 2020.

[77] The reference goes to the well-known criteria laid down by the U.S. Supreme Court in *Daubert v. Merrell Dow Pharmaceuticals*, 509 U.S. 579, concerning: (1) the testability of the method; (2) peer review and publication; (3) the existence of a known or estimable error rate; and (4) the degree of general acceptance of a theory within the relevant scientific community. On this point, see also, for further doctrinal references, DELLA TORRE, *Cripto-attività e giustizia penale*, cit., 340 ff.

[78] As a matter of fact, it is well known that Italian case-law has been inspired for a long time by these criteria to assess the reliability of a given "scientific" method; see, *in primis*, Cass., Sez. IV, 17 September

The need to ensure such a degree of transparency becomes even more compelling when one considers that, in the present context, this safeguard assumes the status of a genuine "meta-right", namely a precondition for the effective exercise of other fundamental guarantees. Indeed, it constitutes an essential safeguard both for verifying the absence of bias within algorithmic systems and for ensuring that such systems do not operate as black boxes.
Nor should it be overlooked that transparency likewise represents an indispensable condition for guaranteeing the effective exercise of a further primary safeguard, now firmly established both at the European[80] and domestic[81] levels as a cornerstone of protection against the risks associated with AI: the existence of "meaningful human oversight" over the outputs generated by algorithmic systems[82]. In practical terms, this entails the adoption of appropriate supervisory measures capable of enabling the human operator not only to understand the functioning of the algorithm, but also effectively to monitor its outputs and to assume genuine responsibility for the resulting decision[83].
It is therefore evident that the scrutiny to be carried out at the stage of admitting AI-based evidence cannot disregard the issue of the reviewability of the system's operation by the human decision-maker. This aspect is of crucial importance: compliance with the principle of the "human in the loop"[84] does not, indeed, consist merely in a formal verification of the proper technical functioning of the system, but requires that the judge be genuinely capable, in the subsequent stages of the evidentiary process, of exercising an autonomous and informed assessment of the material acquired. It follows that, where it is already apparent at the outset that such scrutiny will not realistically be feasible, the process must be halted at the very level of the (in)admissibility of the

2010, n. 43786, in *Dir. pen. proc.*, 2011, 1341, as well as Cass., Sez. IV, 15 December 2025, n. 40243, in *One Legale*.
[79] Cf. QUATTROCOLO, *Prova e intelligenza artificiale*, cit., 480.
[80] Cf. art. 14 Regulation (EU) 2024/1689.
[81] See, in particular, Art. 15, par. 1, Law no. 132/2025, where – with regard to the use of AI in judicial matters – the decision is always reserved to the judge.
[82] Pioneering in this field is the work by UBERTIS, *Intelligenza artificiale, giustizia penale, controllo umano significativo*, in www.sistemapenale.it, 16 November 2020. See also the specific framework of parameters developed to operationalize this principle in ID., *Perizia, prova scientifica e intelligenza artificiale*, cit., 17 f.
[83] Cf., *amplius*, DINACCI, *Intelligenza artificiale tra quantistica matematica e razionalismo critico: la necessaria tutela di approdi euristici*, in *Proc. pen. giust.*, 2022, 1629 f.
[84] Cf., *ex multis*, LETA JONES, *The right to a human in the loop: Political constructions of computer automation and personhood*, in *Social Studies of Science*, 2017, 216 ff.

evidence, given the excessive risk that the output, from being a potentially useful evidentiary tool, may instead become a cognitive trap giving rise to miscarriages of justice[85].

2.2. *The strategic role played by the criteria for the evaluation of evidence.* The relationship between AI evidence and the category of "criteria governing the evaluation of evidence" requires a preliminary conceptual clarification.
In this regard, legal scholarship has put forward various attempts to classify this category of evidence within the general framework of the law of evidence. Among the proposed reconstructions, particular support has been garnered by the view that classifies it as a subcategory of scientific evidence[86], on the assumption that the inference on which it is based cannot be derived from ordinary common knowledge but instead presupposes the use of specialized expertise.
Although this reconstruction captures an important aspect of the phenomenon, it does not appear universally persuasive. Indeed, as we have shown elsewhere in relation to blockchain analysis software, there may be cases in which the inference generated by an algorithmic tool is not grounded in scientific laws *stricto sensu*, but instead rests on empirical regularities or technical maxims of experience[87]. Care must therefore be taken not to automatically extend the aura of scientific validity surrounding such tools to inferences that, in reality, rest upon more fragile epistemological foundations. Otherwise, this would risk reinforcing the already mentioned danger of evidentiary overestimation, which constitutes one of the principal challenges associated with the use of such instruments. It is precisely for these reasons that, from a conceptual standpoint, it appears rather more convincing to classify AI evidence

[85] For some concrete examples of miscarriages caused by these technologies, see SWARNS, *When Artificial Intelligence Gets It Wrong. Unregulated and untested AI technologies have put innocent people at risk of being wrongly convicted*, in *www.innocenceproject.org*, 19 September 2023; SANFORD, *Artificial Intelligence Is Putting Innocent People at Risk of Being Incarcerated*, *ivi*, 14 February 2024.
[86] See, for instance, GIALUZ, *Intelligenza artificiale e diritti fondamentali in ambito probatorio*, in *Giurisdizione penale, intelligenza artificiale ed etica del giudizio*, Milan, 2021, 61 f.; PRESSACCO, *Intelligenza artificiale e ragionamento probatorio nel processo penale*, in *Intelligenza artificiale e processo penale*, ed. by Di Paolo-Pressacco, cit., 119; ROMANÒ, *Intelligenza artificiale come prova scientifica nel processo penale: una sfida tra* machine-generated evidence *e equo processo*, in *Prova scientifica e processo penale*, ed. by Canzio-Lupária Donati, cit., 914 ff.
[87] Cf. DELLA TORRE, *Cripto-attività e giustizia penale*, cit., 353 f.

within the broader category of digital evidence, insofar as it invariably consists of the output of a formalized process carried out by a computational system[88]. However, even more important is the recognition of its inherently circumstantial nature[89]. From this perspective, automated evidence takes the form of a line of reasoning which, proceeding from an established circumstance (for example, a video frame subjected to algorithmic analysis, a blockchain transaction under examination, or a biological trace analyzed through genetic comparison techniques), enables the decision-maker to reach – through the application of inferential rules, whether scientific laws or, indeed, technical maxims of experience – an evidentiary conclusion characterized by a variable degree of persuasiveness. The basic structure of circumstantial reasoning is therefore preserved[90].

It should be noted that the inclusion of AI-based evidence within this genus gives rise to two consequences of considerable importance.

First, it makes it possible to appreciate the epistemic limits characterizing the evidentiary materials in question[91]. Even where an algorithmic tool is highly sophisticated, its cognitive contribution remains anchored to a probabilistic form of reasoning structured as a creative-abductive inference and is therefore inherently exposed to the risk of error. The output generated is incapable of providing "rational certainty"; rather, it is the product of specific training and input data, as well as of operational rules that may be incomplete, imperfect, and context-sensitive. Awareness of these limits provides a significant safeguard against any form of technological determinism, preventing algorithmic evidence from being attributed an apodictic demonstrative value[92]. On the contrary, it must be treated for what it truly is: a circumstantial evidentiary el-

[88] This conclusion is also shared by GIALUZ, *Intelligenza artificiale e diritti fondamentali*, cit., 61 f.; PRESSACCO, *Intelligenza artificiale e ragionamento probatorio*, cit., 119, albeit based on an assumption that is not endorsed here, namely the idea that all forms of computer evidence constitute "scientific" evidence. In our view, such a reconstruction is unpersuasive. The decisive factor is not the digital nature of the instrument employed, but rather the epistemic quality of the inferential rules embedded within it. What ultimately matters is whether the computational process is structured around the application of a scientific law properly so called, or whether it merely operates on the basis of empirical regularities or maxims of experience.

[89] In this regard, see DINACCI, *Intelligenza artificiale*, cit., 1633 f.; PRESSACCO, *Intelligenza artificiale e ragionamento probatorio*, cit., 106.

[90] Cf., for all, TONINI-CONTI, *Il diritto delle prove penali nel sistema accusatorio contemporaneo*, Milan, 2025, 74.

[91] See GIALUZ, *Intelligenza artificiale e diritti fondamentali*, cit., 62.

[92] See, once again, GIALUZ, *Intelligenza artificiale e diritti fondamentali*, cit., 62.

ement whose degree of reliability must be subjected to careful critical scrutiny in light of the conditions of training, the quality of the data employed, and the soundness of the inferential rules on which it is based.

Second – and above all – the classification of AI-based evidence as circumstantial evidence makes it possible to frame the analysis within categories that are well known to Italian criminal procedure, thereby requiring this type of evidence to be assessed in light of the rule of seriousness, precision, and consistency laid down in Article 192(2) of the CCP[93]. This entails a consequence of primary importance: the outputs generated by AI systems cannot, in themselves, attain the status of sole and decisive evidence of guilt, but must instead be corroborated by additional evidentiary elements. And this in full coherence with the decision rule (in its intermediate sense) of the "non-exclusivity of algorithmic decision-making"[94], which may be derived from Article 11 of Directive (EU) 2016/680 and Article 22 of the GDPR, both of which preclude the adoption of fully automated decisions capable of producing adverse legal effects on individuals[95]. This provides concrete confirmation of how one of the most established criteria governing the evaluation of evidence within our legal system may today perform a renewed safeguarding function with respect to cutting-edge forms of technical evidence, operating as a safeguard against the risk that the decision-maker may be misled by an overestimation of algorithmic inferences.

It should be noted that the strategic role assumed by the criteria governing the evaluation of evidence, in light of the growing use of algorithmic tools, finds significant confirmation in Regulation (EU) 2024/1689. Particular reference should be made to the so-called "four-eyes principle" expressly provided for therein, pursuant to which certain outputs generated by high-risk AI systems must be reviewed and independently validated by at least two competent natural persons (Art. 14(5))[96]. The rationale underlying this provision is evident: it introduces an additional epistemic safeguard against the risk of technological determinism, seeking to prevent decision-makers from overestimating the inferential value of algorithmic outputs. In other words, the mechanism of

[93] See PRESSACCO, *Intelligenza artificiale e ragionamento probatorio*, cit., 106.

[94] On this matter, see SIGNORATO, *Il diritto a decisioni penali non basate esclusivamente su trattamenti automatizzati: un nuovo diritto derivante dal rispetto della dignità umana*, in *Riv. dir. proc.*, 2021, 101 ff.

[95] This direction had already been adopted by GIALUZ, *Quando la giustizia penale*, cit., 17.

[96] On this point, see CAMON, *Intelligenza artificiale*, cit., 3019.

dual human verification operates as a form of "institutionalized corroboration", designed to reintroduce a strengthened human-critical review into the decision-making process and thereby neutralize any form of automatic deference to the machine.

This confirms how "negative legal rules governing the assessment of evidence"[97] – namely, rules requiring the corroboration of particularly fragile or sensitive forms of evidence[98] – are capable of playing a central role in contemporary societies, in line with an interpretation that has been widely supported in international literature in recent years[99]. Indeed, when calibrated appropriately, such rules can provide valuable epistemic safeguards, ensuring that criminal adjudication remains rooted in a model of human rationality rather than being superseded by a machine-based model[100].

Nor does the principle of free evaluation of evidence, on which Italian criminal law of evidence is traditionally based, preclude this interpretation[101]. As is well known, the most convincing interpretation of this principle rejects any view of judicial decision-making as an "oracular" exercise likely to lead to arbitrariness. Instead, it conceives judicial discretion as a rational and reviewable activity consistent with the *conviction raisonnée* model[102]. From this point of view, the introduction of well-designed corroboration requirements – espe-

---

[97] Concerning this category, see in particular DANIELE, *Regole di esclusione*, cit., 119 ff., who notes that such rules «prohibit the judge from deriving a certain type of conviction from specific items of evidence, and thus establish a 'negative' rather than a 'positive' evaluation"».

[98] Precisely because negative evidentiary assessment rules structurally entail a logical operation aimed at prohibiting the judge from considering a fact established in the absence of evidentiary elements capable of confirming the reliability of a stigmatized item of evidence, they are also referred to in Anglo-American legal terminology as "corroboration rules": see, for example, MC CONVILLE, *Corroboration and Confessions: The Impact of a Rule Requiring that No Conviction Be Sustained on the Basis of Confession Evidence Alone,* London, 1993.

[99] In this sense, see, for instance, DAMAŠKA, *Evaluation of Evidence. Pre-Modern and Modern Approaches*, Cambridge, 2019, 146 ff. and THAMAN, *Ensuring the factual reliability of criminal convictions: reasoned judgments or a return to formal rules of evidence?*, in *Comparative Criminal Procedure*, ed. By Ross-Thaman, Cheltenham, 2016, 108 ff.

[100] In short, the underlying spirit is to ensure that – to borrow the words of LUPÁRIA DONATI, *Notazioni controintuitive su intelligenza artificiale e libero convincimento*, in *Giurisdizione penale*, cit., 115, themselves echoing F. Carrara – procedural law, if no longer conceived "for gentlemen", may at least remain a body of law designed for human beings rather than for machines.

[101] On this point, the study by NOBILI, *Il principio del libero convincimento del giudice*, Milan, 1974, still remains unsurpassed. More recently, see in particular CARLIZZI, *Libero convincimento e ragionevole dubbio nel processo penale. Storia, prassi, teoria*, San Lazzaro di Savena, 2018.

[102] On this point, see, for all, DAMAŠKA, *Evaluation of Evidence*, cit., 121 ff.

cially for forms of evidence that are structurally fragile or opaque, such as AI-generated evidence – does not only not conflict with the principle of free evaluation but may actually reinforce it. This is because these requirements act as epistemic safety nets, mitigating the risk of arbitrariness and enhancing the transparency and reviewability of judicial decisions by necessitating an additional level of justification during the most critical stages of the reasoning process.

That being said, it should be noted that some scholars have attached such importance to corroboration rules that they have developed a perspective even further removed from the idea of evidentiary evaluation as a "forbidden garden" for the legislator[103]. According to this view, specific regulatory frameworks should be introduced under which, with respect to certain particularly problematic categories of evidence, the judge's reasoned belief alone should not be sufficient to ground a conviction in the absence of corroboration provided by AI systems[104]. What is particularly noteworthy is that, under this approach, the objective would no longer be merely to establish safeguards against AI. Rather, computational tools would themselves become instruments of verification, capable of strengthening the probative value of evidentiary elements that, standing alone, could not legitimately support a decision adverse to the defendant.

It should be acknowledged that this perspective is already appealing from a theoretical standpoint. The introduction of a set of "rules requiring AI corroboration of evidence"[105] would, in fact, revive –albeit in an updated form – an approach associated with some of the leading figures of the Italian Enlightenment, such as Gaetano Filangieri[106]. Indeed, the underlying ambition would be to establish an intermediate evidentiary paradigm, positioned between negative legal proof and free evaluation of evidence, capable of preventing the latter from becoming a «voracious super-logical power»[107]. Moreover, it is precisely because this principle has proved so difficult to govern in practice that one of the leading scholars in the field, Massimo Nobili, argued that «the 'free

[103] The expression is, obviously, taken from FERRUA, *Un giardino proibito per il legislatore: la valutazione delle prove*, in *Quest. giust.*, 1998, 587 ff.

[104] This opinion is adopted by LUPÁRIA DONATI, *Notazioni controintuitive*, cit., 120 f.

[105] The term is taken from LUPÁRIA DONATI, *Notazioni controintuitive*, cit., 121.

[106] Cf. FILANGIERI, *La scienza della legislazione*, t. I, Venice, 1806, 149 f.

[107] The quotation is taken from CORDERO, *Diatribe sul processo accusatorio*, in ID., *Ideologie del processo penale*, Milan, 1966, 229.

conviction' of the judge, whatever beneficial effects it was meant to produce, already did so in the years between Voltaire and Carrara»[108]. It was precisely for this reason that he advocated abandoning a pure conception of the principle in favour of introducing negative rules governing the assessment of evidence[109].
Furthermore, as will be discussed in the final part of this essay[110], this approach also deserves particular attention from a practical perspective. This is especially the case if it is understood as promoting the calibrated use of Bayesian methodologies to assist in the assessment of available evidence – whether AI-generated or not – with a view to reducing the risk of cognitive biases and other distortions in evidentiary reasoning affecting judicial reasoning. As will be argued, and in line with the rationale behind the doctrinal proposal outlined above, introducing a requirement to carry out such verification could enhance the epistemic quality of judicial decision-making in appropriate cases and consequently reduce the risk of error, thus contributing to fundamental objectives of decisional fairness.

2.3. *Standards of proof and AI evidence*. Having explored the safeguarding role played by decision rules in their broader and intermediate meanings in addressing the risks associated with algorithmic evidence, attention can now turn to decision rules in their narrowest sense and, more specifically, to standards of proof.
To this end, attention must first be directed to the principal functions served by evidentiary thresholds: the first is to erect a barrier against arbitrary interferences with individual rights by public authorities[111]. By imposing an epistemic bar that must be cleared before a legal consequence may be imposed, standards of proof constrain the discretion of State authorities and subject the exercise of coercive powers to conditions of evidentiary justification. It is for this reason that they perform a fundamental guarantees-based function, safeguarding rights such as the presumption of innocence, personal liberty, property rights and, in some jurisdictions, even life itself.

---

[108] Cf. NOBILI, *Esiti, errori, arbitrii dietro un'illustre formula: gli ultimi trent'anni,* in *Il libero convincimento del giudice penale. Vecchie e nuove esperienze*, Milan, 2004, 59.
[109] See, once again, NOBILI, *Esiti*, cit., 56.
[110] Cf., *infra*, § 3.1.
[111] Cf. FERRER BELTRÁN-TUZET, *Sulla necessità degli* standard *di prova per la giustificazione delle decisioni giudiziali*, in *Dir. e questioni pubbliche*, 2018, 2, 462.

Secondly, standards of proof perform a crucial justificatory function in relation to decisions taken within procedural settings. By defining the level of proof required, they enable judicial authorities both to make an *ex ante* selection among competing factual hypotheses and to explain *ex post* why, among the hypotheses supported by the evidentiary record, one has been accepted rather than another. It is therefore unsurprising that legal scholarship has long emphasized their close relationship with the reasoning on facts in judicial decisions[112].

Thirdly, standards of proof perform an equally important heuristic function, prompting both public and private actors to seek and produce evidence capable of satisfying the applicable threshold. This effect is particularly significant in criminal proceedings, where it operates as an incentive for investigative authorities to conduct thorough investigations, both quantitatively and qualitatively[113].

Finally, among the crucial functions performed by this legal institute must be included their contribution to the allocation of the risk of erroneous judicial decisions[114]. In order to understand this role, it is useful to distinguish between two categories of error: (a) "false negatives", namely decisions that find an actually true hypothesis not proven; and (b) "false positives", namely decisions that find a false hypothesis proven. It is readily apparent that, as the threshold of a standard is raised, it becomes physiologically more difficult to satisfy it in practice. This has a dual logical consequence: on the one hand, it increases the risk of "false negatives" (i.e. failing to prove a true hypothesis), and on the other hand, it reduces the risk of "false positives" (i.e. finding a false hypothesis proven).

Keeping this feature in mind is essential to understanding why the standards governing final decisions in criminal proceedings are structured as the most demanding among those found in legal systems. The rationale underlying this

---

[112] See for all DAMAŠKA, *Evaluation of Evidence*, cit., 143.

[113] On the principle of "thoroughness of investigations", as first emphasized by the Constitutional Court, in Judgment No. 88 of 15 February 1991 and, most recently, further reinforced in Judgment No. 58 of 27 April 2026, see MARZADURI, *Qualche considerazione sui rapporti tra principio di obbligatorietà dell'azione penale e completezza delle indagini preliminari*, in *www.sistemapenale.it*, 14 May 2020; SIRACUSANO, *La completezza delle indagini nel processo penale*, Turin, 2005; VALENTINI, *La completezza delle indagini, tra obbligo costituzionale e (costanti) elusioni della prassi*, in *Arch. pen. web*, 2019, 3.

[114] For a deeper analysis of this theme, see LAUDAN, *The Law's Flaws. Rethinking Trials and Errors?*, Milton Keynes, 2016, 103 ff.; STEIN, *Foundations of Evidence Law*, Oxford, 2005, 118 ff.

approach lies in the fact that the interests at stake in criminal proceedings are among the most fundamental. For this reason, they are designed not primarily to minimize the overall incidence of fact-finding errors, but rather to reduce the risk of the most serious type of error, namely the conviction of an innocent person (a false positive). From this perspective, the *quantum* of proof required for a criminal conviction is inextricably linked to the presumption of innocence, understood in its function as a strict rule of adjudication, that is, a norm allocating the risk of non-persuasion to the party bearing the burden of proof.

Against this background, the importance of standards of proof becomes even more apparent when considering the risks associated with algorithmic evidence. In this context, the functions discussed above acquire renewed significance, as they are closely connected to the requirement of ensuring meaningful human control over algorithmic inferences and the risks associated with them.

As is obvious, this is particularly true of the criminal standard of proof beyond a reasonable doubt under Article 533 CCP. Precisely because it serves as a safeguard against the risk of wrongful convictions, the assessment it requires cannot, unlike the evaluation carried out at the stage of evidentiary admissibility, be exhausted merely by reference to the abstract transparency of the model or to its hypothetical error rate. Rather, it must take the form of a concrete and multi-layered inquiry. At a minimum, such an inquiry should ascertain: (a) whether, in the specific case, the algorithm has operated consistently with its underlying premises and without opaque inferential leaps ("scrutiny of inferential functioning")[115]; (b) whether the output obtained is explainable in light of the particular conditions of use and has not been adversely affected by contextual factors capable of undermining its reliability ("scrutiny of contextual reliability")[116]; (c) whether the output coheres with the overall body

[115] Consider, for instance, a biometric recognition system which, despite being characterized by a known and acceptable error rate, signals a match in a specific case without making it possible to determine which features generated that result or how they were weighted by the model. In such a situation, the system may be reliable in the abstract, yet the decision-maker lacks any meaningful possibility of verifying the correctness of the inferential process in the concrete case.

[116] Consider a facial recognition system trained predominantly on images captured under favourable lighting conditions and subsequently applied to a night-time frame or a low-resolution image. Even assuming that the system operates in accordance with its design specifications, the output may nevertheless prove unreliable in light of the specific conditions of use.

of available evidence and with the scientific and/or experiential generalizations governing the functioning of the world ("scrutiny of overall epistemic soundness")[117].

In all such cases, where significant uncertainty or unresolved inconsistencies persist regarding the output of an algorithm relied upon as incriminating evidence, a form of reasonable doubt "internal" to the prosecution's reconstruction inevitably arises. According to well-established case law, this notion encompasses situations revealing either internal contradictions or – as in the present context – the explanatory inadequacy of the prosecutorial hypothesis[118]. Put differently, where, following the evidentiary assessment, it remains impossible to ascertain whether and how the algorithm actually produced the result in question, or whether significant margins of error remain insufficiently controlled, the resulting uncertainty necessarily affects the reliability of the inference drawn from the algorithmic output itself.

It is equally clear that this line of reasoning must be expressly and specifically reflected in the statement of reasons accompanying the judgment. This follows from the fact that the safeguards outlined above are complemented by the set of rules governing the rational justification of judicial decisions, which give effect to Article 111(6) of the Constitution. More specifically, an important additional layer of protection arises from the interaction between the duty of reasoned and dialogical decision-making incumbent upon the judge (Articles 192(1) and 546 CCP) and the system of remedies available to the parties, both on the merits (appeal under Articles 593 ff. CCP) and on points of law (appeal to the Supreme Court under Articles 606 ff. CCP). Taken together, these provisions perform a crucial closing-guarantee function within the legal system, as they help ensure respect for the fundamentally human epistemic purpose of criminal proceedings[119]. Following the admission and assessment of technical evidence, the judge must therefore explain, in clear, ra-

[117] Consider an algorithmic output attributing a subject to a particular scene, despite being incompatible with other available evidentiary elements, such as geolocation data or consistent witness testimony placing that individual elsewhere. In such a case, even if the result is technically intelligible, it cannot be coherently integrated into the overall evidentiary framework. This inconsistency may in turn indicate either data manipulation or a malfunction of the system, thereby undermining the reliability of the output in the concrete case.

[118] Cf., *ex multis*, Cass., Sec. I, 24 October 2011, n. 41110, in *Ced. Cass.*, n. 251507, as well as, more recently, Cass., Sec. V, 2 October 2023, n. 39777, in *One Legale*.

[119] On the endorsement of this objective, see MAZZA, *Il processo che verrà: dal cognitivismo garantista al decisionismo efficientista*, in *tArch. pen. web.*, 2022, 2, 3.

tional, and humanly intelligible terms, how it was possible to move from an incriminating circumstance to the fact to be proven through reliance on an inherently fallible instrument such as AI software, thereby demonstrating that meaningful control over the algorithmic inference has in fact been exercised. The importance of this guarantee becomes even more apparent when one considers that deficiencies in the reasoning process may subsequently be challenged through powerful remedial mechanisms, including appeal and, where appropriate, review by the Supreme Court.

If such a procedural model applies to the standard of proof beyond a reasonable doubt, extending a similarly demanding approach to intermediate thresholds may, at first sight, appear more problematic. This is true, in particular, of standards such as the "serious indications of guilt", required for the application of personal precautionary measures under Article 273 CCP, and the "reasonable prospect of conviction", governing both the decision between prosecution and dismissal (Article 408 CCP) and that between committal for trial and a judgment of no grounds to proceed (Articles 425 and 554-ter CCP), especially given that they have traditionally been understood as requiring a lower degree of evidentiary support than that required for the final decision under Article 533 CCP[120].

However, in our view, an interpretation consistent with the safeguarding function performed by these evidential thresholds likewise requires that they be applied with equal rigor in the context under consideration.

With regard to the criterion of "serious indications of guilt", it should first be recalled that the prevailing case law identifies a twofold reduction in the epistemic demands of precautionary proceedings as compared with the beyond a reasonable doubt standard[121]. First, it excludes the need for a univocal reconstruction of the facts, considering the imposition of coercive measures compatible with the coexistence of alternative hypotheses supported by the same evidentiary basis (thus allowing for the presence of doubts external to the prosecution's reconstruction)[122]. Second, it accepts that the evidentiary record at the precautionary stage is inherently incomplete and may be supplemented as the proceedings progress. It is also for this reason that the prevailing case

[120] In this regard, see ORLANDI, *Discrezionalità e regole di giudizio nel processo penale*, cit., 29 ff.

[121] For a difference between the two standards, cf., for instance, Cass., Sec. I, 7 January 2025, No. 323, in *One Legale*.

[122] Cf. Cass., Sec. III, 7 May 2024, n. 17839, in *One Legale*.

law excludes, at the stage of applying liberty-restricting measures, the requirement that indications be precise and concordant within the meaning of Article 192(2) CCP, considering the sole requirement of seriousness sufficient for the imposition of the measure[123]. However, it is precisely the requirement of seriousness that, in our view, cannot be regarded as satisfied unless AI-based evidence has first undergone the threefold scrutiny of inferential functioning, contextual reliability, and overall epistemic soundness. Absent a positive outcome of such scrutiny, it is indeed impossible to determine whether the result produced falls within the system's margins of reliability or, conversely, within one of its zones of error, nor whether it has been affected by contextual factors capable of undermining its trustworthiness. Nor could it be otherwise from a normative perspective: at the precautionary stage, fundamental individual rights are at stake, and these cannot legitimately be restricted except on the basis of sufficiently reliable and verifiable evidentiary elements[124].

In our view, this conclusion applies with even greater force to the standard of the "reasonable prospect of conviction". This follows from the fact that constitutional case law has, particularly in recent years, adopted a notably de-

[123] See Cass., Sec. V, 15 January 2025, n. 1850, in *One Legale*.

[124] In light of these considerations, the practice that has developed in Italy with regard to the use of the automated facial recognition system S.A.R.I. appears particularly objectionable (see DEMARTIS, *I sistemi automatici di riconoscimento facciale nel procedimento penale. Tra possibilità di impiego e limiti ordinamentali*, Milan, 2025, 106 ff.; SACCHETTO, *Tecnologie di riconoscimento facciale e procedimento penale. Indagine sui fondamenti e sui limiti dell'impiego della biometria moderna*, Turin, 2025, 154 ff.). Indeed, although the lack of adequate information has repeatedly been denounced both with regard to the error rate and the safeguards designed to prevent discriminatory uses (on this point, see DELLA TORRE, *Tecnologie di riconoscimento facciale e procedimento penale*, in *Riv. it. dir. proc. pen.*, 2022, 1082), this tool is now routinely employed in investigative practice, in precautionary proceedings, and even at trial stage, amid substantial judicial indifference (see, most recently, Cass., Sec. II, 20 March 2025, No. 18099, in *One Legale*, which rejected the defendant's claim that it was illogical to attribute evidentiary value to a mere 55.2% match between images of an individual captured by CCTV and his own mugshot, on the ground that, in the reasoning of the trial court, S.A.R.I. had not been used as sole and decisive evidence. In the context of precautionary measures, see Cass., Sec. IV, 13 July 2023, No. 39551, in *One Legale*, which held the system admissible where it forms part of a broader evidentiary framework. For a broader analysis of case law, see BELVINI, *Artificial Intelligence*, cit., 170). Such an approach is difficult to endorse, particularly because, as several scholars have observed, S.A.R.I. should have been regarded as inadmissible from the outset in the absence of a clear domestic regulatory framework capable of ensuring compliance with the safeguards required by Article 189 CCP (on this point, *inter alia*, BELVINI, *ibid.*, 175; GIALUZ, *Intelligenza artificiale*, cit., p. 64; TRIGGIANI, *L'utilizzo investigativo e probatorio dei sistemi automatici di riconoscimento facciale dopo l'AI Act e la l. n. 132/2025: dubbi e perplessità*, in *Giudizio penale e intelligenza artificiale*, cit., 330).

manding approach in this area. On the one hand, it has emphasized the connection – albeit in a prognostic sense – between this standard and the beyond a reasonable doubt threshold[125]. On the other hand, it has been underlined that even the mere exercise of prosecutorial powers, insofar as it affects constitutionally protected interests, must be guided by requirements of thoroughness and accuracy, as well as by particular caution and responsibility[126]. Against this background, it becomes apparent that the standard of "reasonable prospect of conviction" cannot be regarded as satisfied where the algorithmic elements underpinning the decision to prosecute are not characterized by a minimum degree of transparency and verifiability. To hold otherwise would create a cognitive deficit capable of undermining the very possibility of rational scrutiny of the evidentiary inference.

The foregoing considerations lead to the conclusion that AI-based evidentiary elements must already be accompanied, at the investigative stage, by a sufficient body of information enabling meaningful human review of their inferential functioning, reliability, and overall epistemic soundness in light of the available data concerning the algorithmic output. Conversely, where this threshold is not met and significant areas of opacity persist even upon a *prima facie* assessment conducted before the activation of adversarial proceedings, the evidence should be regarded as inherently unreliable and therefore incapable of justifying either the progression of the proceedings or the adoption of precautionary measures.

In short, these considerations demonstrate that standards of proof are by no means a mere "relic" of the past, as has occasionally been suggested[127]. On the contrary, they constitute a normative category endowed with renewed vitality, capable of performing a crucial safeguarding function against the misuse of computational tools.

---

[125] In this regard, see Constitutional Court, 15 October 2024, n. 179 and, more recently, Corte cost., 27 April 2026, n. 58.

[126] See also Constitutional Court, Judgment No. 58 of 27 April 2026, which, starting from the serious adverse effects that the mere pendency of criminal proceedings may have on an individual's professional and social life, expressly departed from its earlier Judgment, No. 88 of 15 February 1991, holding that it is no longer sustainable to infer from Article 112 of the Constitution a general *favor actionis*, according to which, in cases of doubt, prosecution should be pursued rather than withheld.

[127] *Contra*, see NIEVA FENOLL, *Carga de la prueba y estándares de prueba: dos reminiscencias del pasado*, in *Estudios de Derecho*, 2020, 170, 119 ff.

3. *The algorithm as a tool for the evaluation of evidence and decision-making: the Bayesian approach to the law of evidence*. Having explored the challenges posed by algorithmic evidence, the analysis may now turn to the second part of this inquiry, namely the role of Bayesian methodologies in legal evidentiary reasoning[128].

As a preliminary observation, it is worth recalling that the debate on the possibility of representing the inferential processes of evidentiary reasoning through formal tools, although rooted in a long-standing intellectual tradition[129], gained significant momentum within the framework of the so-called New Evidence Scholarship[130]. Initially developed in the Anglo-American world from the 1970s onwards, this movement gradually expanded and acquired global influence.

Despite the diversity of positions adopted, many scholars associated with this school of thought have gradually converged around three fundamental ideas: (a) first, the notion that legal evidentiary reasoning, being non-deductive and predominantly inductive and abductive in nature, operates under conditions of uncertainty and therefore possesses an intrinsically probabilistic dimension[131]; (b) second, the postulate that propositions relating to individual cases may be expressed in terms of "subjective probability"[132], mathematically formalized and understood as a measure of the rational degree of belief attribut-

[128] For a detailed presentation of the Bayeasian approach, cf., in Italian: CHERUBINI, *Guida al ragionamento probatorio con le reti* bayesiane, Milan, 2024; GARBOLINO, *Probabilità e logica della prova*, Milan, 2014; MURA, *Teorema di Bayes e valutazione della prova*, in *Cass. Pen.*, 2004, 1808 ff.; TILLERS-GREEN (ed. by), *L'inferenza probabilistica nel diritto delle prove. Usi e limiti del* bayesianesimo, Milan, 2003. For a discussion in the criminal procedure literature, see, in particular: CATALANO, *Logica della prova*, statistical evidence *e applicazione della teoria delle probabilità nel processo penale*, in *Dir. pen. cont., Riv. trim.*, 2013, 4, 132 ff.; EAD., *Prova indiziaria,* probabilistic evidence *e modelli matematici di valutazione*, in *Riv. dir. proc.*, 1996, 518 ff.; FERRUA, *La prova nel processo penale*, in *Riv. Bras. Dir. Proc. pen.*, 2018, 4, 116 f.; LUPÁRIA DONATI, Trials by probabilities, cit., 95 ff.; TRAPELLA, *Introduzione ad una ricerca su indagini preliminari, intelligenza artificiale e* bias *del ragionamento*, in *Arch. pen. web*, 2025, 3, 7 f.; UBERTIS, *Profili di epistemologia giudiziaria*, Milan, 2021, 180 f.

[129] In this regard, see GARBOLINO, *Probabilità e logica della prova*, cit., 435 ff.

[130] On this point, see LEMPERT, *La nuova dottrina delle prove: un'analisi della formazione della prova concludente*, in *L'inferenza probabilistica nel diritto delle prove. Usi e limiti del* bayesianesimo, cit., 91 ff. and TWINING, *The New Evidence Scholarship*, in *Cardozo Law Review*, 1991, 2-3, 295 ff.

[131] This starting point is, moreover, shared also by the most influential non-Bayesian theories of the logic of legal evidence. What differs, rather, is the conception of probability deemed appropriate to represent evidentiary inferences: for a clear classification of the heterogeneous approaches on this issue, see TUZET, *La prova ragionata*, cit., 219 ff.

[132] On this point, cf. FERRER BELTRÁN, *La valutazione razionale della prova*, Milan, 2007, 107 ff.

ed to their truth; (c) third, the idea that Bayes' theorem provides the most appropriate framework for developing such reasoning within the legal domain, insofar as it is capable of structuring the inferential process, making its premises explicit and rendering its logical steps more amenable to scrutiny[133].

For present purposes, it is sufficient to focus on the core element of this model, namely the "Bayesian rule", which may be defined as the «logical-probabilistic function describing the correct procedure for revising confidence in a hypothesis in light of the available evidence»[134].

To illustrate how the model operates, a simple example may be useful[135]. Assume that one wishes to estimate the probability that it will rain on a given day, having at one's disposal meteorological statistics indicating that, in a certain geographical area, rainfall occurs on more than half of the days of the year. This information provides an initial estimate of the prior probability of the hypothesis that it will rain. Now consider two additional pieces of evidence: a completely clear sky and a weather forecast excluding precipitation. Intuitively, both elements are difficult to reconcile with the hypothesis that rain is imminent and are instead considerably more consistent with the alternative hypothesis, namely that it will not rain, since they are much more likely to occur in that scenario. As a result, incorporating this new information leads to a downward revision of the posterior probability of the original hypothesis relative to the initial estimate.

This example captures the essential logic of the Bayesian method, which consists in the progressive updating of prior beliefs concerning a given hypothesis in light of newly acquired evidence[136]. The value of the model lies precisely in

---

[133] For scholarly support for the use of Bayesian methods in legal decision-making, see, *inter alia*, DAHLMAN, *De-Biasing Legal Fact-Finders With Bayesian Thinking*, in *Topics in Cognitive Science*, 2020, 12, 1115 ff.; EGGLESTON, *Prova, conclusione probatoria e probabilità*, Milan, 2004, 169 ff.; FINKELSTEIN-FAIRLEY, *A Bayesian approach to Identification Evidence,* in *Harvard Law Review*, 1970, 83(3), 489 ff.; FRIEDMAN, *Answering the Bayesioskeptical Challenge*, in *The International Journal of Evidence and Proof*, 1997, 1, 276 ff.; GARBOLINO, *Probabilità e logica della prova*, cit., *passim*; LEMPERT, *La nuova dottrina delle prove: un'analisi della formazione della prova concludente*, in *L'inferenza probabilistica nel diritto delle prove. Usi e limiti del* bayesianesimo, cit., 91 ff.; TARONI-BOZZA-VUILLE, *La probabilità come strumento per una coerente valutazione della prova scientifica*, in *Prova scientifica e processo penale*, ed. by Canzio-Lupária Donati, cit., 33 ff.

[134] The quotation is taken from CHERUBINI, *Trappole cognitive nel ragionamento giudiziario*, in *Cass. Pen.*, 2007, 3, 1363.

[135] The example is drawn from MACKOR, *Bayesian Modelling of Criminal Cases as a Whole: a Philosophical Reflection on Dutch Case Law*, in *Quaestio facti*, 2026, 10, 365.

[136] Cf. TARONI-BOZZA-VUILLE, *La probabilità*, cit., 36.

its ability to formalize this inferential process, thereby making it more transparent, structured, and open to rational scrutiny.
In particular, the version of Bayes' theorem most frequently used in legal reasoning is the odds form[137], according to which:

$$P(H \mid E) / P(\text{non-}H \mid E) = P(H) / P(\text{non-}H) \times P(E \mid H) / P(E \mid \text{non-}H)$$[138].

More specifically, the formula is composed of two fundamental elements.
The first is the ratio P(H) / P(non-H), which expresses the relationship between the prior probabilities of the competing hypotheses, namely the degree of belief initially attributed to H and non-H before the available evidence is taken into account (in the example above, the statistical frequency of rainfall in the relevant geographical area). This component requires consideration of the broader context in which the hypothesis is situated and plays a decisive role in determining the impact of subsequent evidence. Even highly probative evidence may have only limited effects when associated with an initially unlikely hypothesis, whereas the same evidence may prove decisive when it reinforces a hypothesis that is already supported by a rational evidentiary basis[139].
The second component is the ratio P(E|H) / P(E|non-H), commonly referred to as the "likelihood ratio". As legal scholars have observed, its meaning may be more readily understood if it is conceived as the ratio between the probability of a "true positive" – that is, the observation of the evidentiary element when hypothesis H is true – and the probability of a "false positive", namely the observation of the same element when the hypothesis is false[140].
For this reason, the likelihood ratio is commonly described in the literature on legal evidence as a measure of the "strength of evidence"[141], since it indicates the extent to which a given item of evidence discriminates between competing hypotheses. More specifically, where the likelihood ratio exceeds

[137] A clear mathematical analysis of this model can be found in GARBOLINO, *Probabilità e logica della prova*, cit., 98 ff.
[138] In this expression: (1) "P" stays for probability; (2) "H" represents the hypothesis under consideration; (3) "non-H" represents its negation, to be understood as an alternative hypothesis incompatible with "H"; (4) "E" stays for the set of available evidentiary elements; and (5) the symbol "|" expresses a conditional relationship and is read as "given that".
[139] Cf. CHERUBINI, *Trappole cognitive nel ragionamento giudiziario*, cit., 1364.
[140] See, once again, CHERUBINI, *Trappole cognitive nel ragionamento giudiziario*, cit., 1363.
[141] See DAHLMAN, *De-Biasing Legal Fact-Finders*, cit., 1117.

1, the evidence increases the posterior probability of the hypothesis under consideration (in our example, H). Conversely, where it is lower than 1, the evidence supports the alternative hypothesis (non-H). Finally, where it is equal to 1, the evidence is neutral, having no effect on the comparative assessment of the competing hypotheses.
The interaction between these two components makes it possible to determine the relative posterior probabilities of competing hypotheses in light of the available evidence and, consequently, to update one's initial degree of belief in a rational manner according to the probative force of the evidence. In the example above, this corresponds to revising the assessment of the competing hypotheses "it will rain" and "it will not rain" in light of the observed weather conditions and the available forecasts.
Considering its general structure, it is unsurprising that the Bayesian model has been regarded as potentially useful, albeit with the necessary adaptations, in the specific context of legal evidentiary reasoning. Indeed, it provides a rigorous methodology for assessing, within a logically transparent and reviewable framework, how a given item of evidence – or a plurality of items progressively acquired during the proceedings – affects the prior probability of a hypothesis, thereby strengthening or weakening the degree of rational belief that may be placed in it. Moreover, the method imposes an inherently comparative structure on evidentiary reasoning, requiring the hypothesis under consideration to be assessed against competing explanatory hypotheses. In this respect, it appears particularly consistent with the procedural requirement that alternative factual reconstructions be identified and critically examined.
What is particularly relevant for present purposes is that, in recent decades, the merely "analogical" use of the Bayesian method has been accompanied by the progressive development of computational applications based on Bayesian networks[142]. These consist of probabilistic graphical models, typically implemented through specialized software, capable of representing relationships of probabilistic dependence among different relevant variables (the so-called nodes), corresponding to factual hypotheses and evidentiary elements capable of affecting them. From this perspective, such tools fall within the broad notion of artificial intelligence adopted by the AI Regulation, insofar as

[142] For a deeper analysis of this instrument, cf. TARONI-ATIKEN-GARBOLINO-BIEDERMANN, *Bayesian Networks and Probabilistic Inference in Forensic Science*, Chichester, 2006 e LAGNADO, *Explaining the Evidence*, Cambridge, 2022, 36 ff.

they are computational systems that, through the application of algorithms, assist and inform human decision-making processes. More specifically, Bayesian networks enable the user to structure and analyse complex inferential scenarios while explicitly and transparently estimating the impact of available evidence on the relative probability of competing factual reconstructions[143].
It should, however, be noted that, in criminal proceedings, the propositions relevant to evidentiary reasoning are traditionally classified according to different hierarchical levels, reflecting the distinct objects of the fact-finding process[144]; and it is precisely in light of this distinction that the role and potential usefulness of the Bayesian model must be assessed. In particular, three fundamental levels are generally identified[145]: (a) the *source level*, concerning propositions relating to the origin of a given material trace; (b) the *activity level*, referring to propositions concerning a specific human activity or event; and (c) the *crime level*, encompassing propositions relating to the commission of the criminal offence.
It is readily apparent that the transition from one level to another entails a progressive increase in inferential complexity, since each raises distinct theoretical and practical issues that require separate examination in the following sections.

3.1. *The Bayesian approach across different levels of evidentiary analysis: the source level and the activity level.* The level at which the Bayesian method has achieved its most significant practical success is undoubtedly the *source level*. In this context, a broad consensus has gradually emerged within the forensic science community that Bayesian reasoning provides the most appropriate framework through which experts – whether court-appointed or party-appointed – can rigorously express the inferential contribution of scientific ev-

[143] For a concrete example of the use of Bayesian networks to a famous Italian case, see CHERUBINI, *Guida al ragionamento probatorio*, cit., 409 ff. Instead, for an interesting Dutch case, see FENTON-NEIL-YET-LAGNADO, *Analyzing the Simonshaven case using Bayesian networks*, in *Topics in Cognitive Science*, 2020, vol. 12, 1092 ff.
[144] On this point, see *European Network of Forensic Science Institutes* (ENFSI), *Guideline for evaluative reporting in forensic science*, in *https://enfsi.eu/wp-content/uploads/2016/09/m1_guideline.pdf*, 27 September 2016, version 3.0, 24. See also COOK-EVETT-JACKSON-JONES-LAMBERT, *A hierarchy of propositions: deciding which level to address in casework*, in *Science & Justice*, 1998, 4, 231 ff.
[145] For a more analytic classification, which includes also the "*sub-source level*" and even the "*sub-sub source level*", cf. SALLAVACI, *Algorithms on Trial: Does evaluative probabilistic reporting of forensic evidence infringe the presumption of innocence?*, in *Forensic Science International: Synergy*, 2025, 9.

idence to factual hypotheses concerning the attribution of the origin of material traces relevant to criminal proceedings. This approach is now explicitly endorsed by several leading reference documents in the field, including the recent ISO 21043-4:2025 standard[146] and the ENFSI guidelines[147], both of which identify the likelihood ratio as the preferred parameter for communicating the evidential weight of scientific and technical findings.

To illustrate this approach more concretely, consider a simple example[148]. Suppose that an expert compares a shoeprint found at the crime scene with the shoe of a suspect and formulates two competing hypotheses: (H1) the shoeprint was produced by the suspect's shoe; and (H2) the shoeprint was produced by another shoe sharing similar characteristics. Under the Bayesian framework, the outcome of the analysis is expressed through a likelihood ratio which, as noted above, measures the extent to which the evidence supports one hypothesis rather than the competing alternative. This may be communicated in quantitative terms by stating, for example, that the observed results are 1,000 times more likely if the shoeprint originated from the suspect's shoe than if it originated from another similar shoe. At the same time, it is widely recognized that the interpretation of complex numerical values is often difficult for human decision-makers, particularly where they lack specific training in probabilistic reasoning. For this reason, the ENFSI guidelines also provide qualitative verbal scales corresponding to different ranges of likelihood ratios, thereby facilitating communication and understanding of evidential weight[149].

However, the advantages of an approach grounded in Bayesian reasoning are not confined to a more precise delineation of the respective roles of expert and judge, thereby avoiding undue overlap between their functions. They also emerge from its capacity to structure the assessment of evidence according to logically coherent criteria. As emphasized in a vast body of scholarly literature, Bayesian methodology constitutes a particularly valuable instrument for preventing errors in the interpretation and use of scientific and technical evi-

[146] For a comment, see BERGER, *Finally a really forensic worldwide standard: ISO 21043 Forensic sciences, Part 4, Interpretation*, in *Forensic Science International: Synergy*, 2025, 10.

[147] *ENFSI, Guideline for evaluative reporting*, cit., 6.

[148] This example is also taken from MACKOR, *Bayesian Modelling*, cit., 366.

[149] Therefore, for instance, a value included between 2 and 10 is described as a "*weak support*"; between 10 and 100 as a "*moderate support*"; between 100 and 1000 as a "*moderately strong support*"; between 1000 and 10.000 as a "*strong support*", and so on. *ENFSI, Guideline for evaluative reporting*, cit., 17.

dence[150]. Among these, the most widely discussed is the so-called "prosecutor's fallacy"[151].

The point may be illustrated by reference to a case decided by the Swedish Supreme Court in 2001 concerning a burglary at a watch shop[152]. On that occasion, one of the perpetrators injured himself on shards of broken glass at the crime scene, leaving blood traces compatible with the defendant's DNA profile. At trial, the forensic expert testified that the probability of a random match – that is, the probability that a person unrelated to the offence might nonetheless possess a genetic profile compatible with the traces recovered at the scene, thereby generating a false positive – was extremely low. The Court, however, fell precisely into the prosecutor's fallacy, assuming that the extremely low statistical probability of such a random match was equivalent to certainty that the defendant was one of the perpetrators of the offence[153].

In technical terms, this error consists in a reversal of the conditional probability, whereby the very low probability of observing the evidence on the assumption that the defendant is innocent is confused with the probability that the defendant is innocent given that same evidence[154]. In doing so, however, one overlooks the role played by prior probabilities and, more generally, by the overall evidentiary context within which the scientific evidence is situated.

---

[150] Cf., *ex multis*, also for further doctrinal references, AITKEN-TARONI, *Statistics and the Evaluation of Evidence*, 2ª ed., Chichester, 2004, 78 ff.; DAHLMAN, *A systematic account of probabilistic fallacies in legal fact-finding*, in *The International Journal of Evidence & Proof*, 2024, 28(1), 45 ff.; ID., *De-Biasing Legal Fact-Finders*, cit., 1121 f.; FENTON-NEIL, *Avoiding probabilistic reasoning fallacies in legal practice using Bayesian networks*, in *Australasian Journal of Legal Philosophy*, 2011, 114 ff.

[151] The label derives from THOMPSON-SCHUMANN, *Interpretation of statistical evidence in criminal trials: The prosecutor's fallacy and the defence attorney's fallacy*, in *Law and Human Behavior*, 1987, 11(3), 167 ff. The fallacy consists in equating the probability of a "false positive" – namely P(E|non-H) – with the inverse probability that the proposition is false in light of the observed evidence, namely P(non-H|E). In other words, it involves confusing the probability of observing a given item of evidence on the assumption that a proposition is false – a factor relevant to the likelihood ratio – with the probability that the proposition is false given the evidence acquired, which instead pertains to posterior probability. On this point, see DAHLMAN, *A systematic account*, cit., 52.

[152] Swedish Supreme Court, 12 December 2003, B 438-03, in *NJA* 2003, 519.

[153] On this point, cf. DAHLMAN, *A systematic account*, cit., 52, to whom reference is made also for further bibliographic references on this case.

[154] Not coincidentally, this fallacy falls within the broader genus of errors of "conditional transposition", on which see, *inter alia*, DAHLMAN, *A systematic account*, cit., 50 ff. See also MACKOR, *Bayesian Modelling*, cit., 367 f., who offers a particularly effective illustration of the distinction between these two probabilities: while the conditional probability that an animal has four legs, given that it is a cow, is very high, the conditional probability that an animal is a cow, given that it has four legs, is comparatively very low.

The crucial point is that even highly discriminating scientific evidence is not, in itself, sufficient to justify rational certainty as to guilt where the prosecution hypothesis is characterized by a particularly low prior probability or lacks adequate support from the remaining body of evidence[155]. This example illustrates both the delicacy of probabilistic reasoning in judicial fact-finding and the significant difficulties that even professional judges, no less than experts themselves, may encounter in applying it correctly[156].

It is precisely in this respect that one of the principal advantages of the Bayesian approach becomes apparent, namely its ability to prevent such inferential distortions by avoiding the automatic equation of the statistical strength of evidence with the probability of guilt. The likelihood ratio measures only the extent to which a given item of evidence discriminates between competing hypotheses and, by itself, does not warrant any conclusion regarding the ultimate factual determination. That assessment must instead be carried out by the judge considering the overall evidentiary record and the prior probability of the competing hypotheses[157]. Although Bayesian methodology cannot eliminate every possible flaw in evidentiary reasoning, this corrective function is of particular significance, especially given that distorted uses of statistics in criminal proceedings – such as the one just described – are far from infrequent and rank among the most insidious causes of miscarriage of justice[158].

Against this background, particular attention should be paid to the proposal – already mentioned above[159] – to introduce an evidentiary assessment criterion

---

[155] In this regard, cf. the judgement delivered by the Tribunal of Milan, sez. G.I.P., sent. 18 June 2015, giud. Gennari, in *Dir. pen. cont.*, 21 October 2015, which will be further examined in the next paragraph.

[156] See SALLAVACI, *Algorithms on Trial*, cit., 7. On the difficulties associated with the correct interpretation and use of statistical information, see the well-known study by HUFF, *Mentire con le statistiche*, Pescara, 2007; see also, more generally, TVERSKY-KAHNEMAN, *Il giudizio in condizioni di incertezza: euristiche e* bias, in Kahneman-Slovic-Tversky, *Decidere nell'incertezza*, Milan, 2024, 11 ff. It is precisely for this reason that some legal systems – notably the Dutch one – have, in recent years, institutionalized the role of "forensic advisers", whose task is to assist judges in the assessment of complex evidentiary material, thereby helping to prevent inferential errors and logical fallacies such as those discussed above. On this point, see. MEEUWISSEN (et al.), *Forensic advisers working for all district courts and courts of appeal in the Netherlands: An overview and discussion*, in *Journal of Forensic Science*, 2024, 69(1), 182 ff.

[157] In this regard, see GARBOLINO, *Probabilità e logica della prova*, cit., 107.

[158] Consider some paradigmatic events such as the British case of Sally Clark and the Dutch case of Lucia de Berk; on these cases, see GARBOLINO, *Probabilità e logica della prova*, cit., 387 ff.

[159] Cf. the proposal by LUPÁRIA DONATI, *Notazioni controintuitive*, cit., 120 f. analyzed *supra* § 2.2.

requiring, at least for certain categories of complex evidence, a structured verification based on Bayesian reasoning as a means of testing the robustness of the inferential path followed by the decision-maker. Such a solution would make it possible to institutionalize a form of rational scrutiny over evidentiary inferences, thereby reducing the risk of logical fallacies capable of undermining the epistemic quality of criminal fact-finding.

A different picture emerges, however, once the analysis moves from the *source level* to the *activity level*[160]. As the ENFSI guidelines emphasize, activity-level propositions become relevant whenever the interpretation of evidence requires specialized knowledge concerning the circumstances in which it was generated. In particular, factors such as mechanisms of trace transfer, persistence over time, and background levels of material may substantially influence the meaning attributed to scientific findings in relation to the activities under consideration[161]. Examples include trace evidence such as fibers, glass fragments, or gunshot residues, as well as small quantities of DNA, narcotics, or explosives. In such cases, the mere presence of a trace may be misleading if it is not properly contextualized.

Against this background, it becomes apparent that evidentiary assessment at the activity level is significantly more complex and, for that very reason, more exposed to uncertainty[162]. In fact, compared with source-level propositions, an additional inferential step is required: the expert must relate the scientific findings to the possible human activities that may have generated them. This is a particularly delicate task, insofar as it involves evaluative judgments that are not entirely determined by empirical data. It is therefore unsurprising that part of the literature has questioned the appropriateness of extending probabilistic analysis to this level[163].

Nevertheless, even this limitation has, in some instances, been overcome in Italian judicial practice using so-called Bayesian networks[164]. However, it is evident that the use of such tools introduces additional layers of complexity.

---

[160] For a clear distinction between the two levels, cf. TARONI-BOZZA, *Il significato delle evidenze: l'importanza dello scienziato forense nella prevenzione dell'errore*, in *L'errore giudiziario*, ed. by Lupária Donati, Milan, 2021, 687 f.

[161] *ENFSI, Guideline for evaluative reporting*, cit., 11.

[162] On this point, cf. GENNARI, *I giudici sono pronti per l'*Activity Level, in *www.sistemapenale.it*, 8 November 2023.

[163] See GENNARI, *I giudici sono pronti per l'Activity Level*, cit.

[164] As reported by GENNARI, *I giudici sono pronti per l'Activity Level*, cit.

First, the question arises as to who should be responsible for constructing the model. Secondly, it becomes necessary to determine which variables ought to be included and according to what criteria, particularly where sufficiently robust statistical regularities are lacking to ground the inferential process[165]. Finally, it must be remembered that the outcome of the analysis is inherently dependent on the information available to the model and may therefore vary significantly as that information changes[166].

At the same time, it should not be overlooked that the algorithmic approach also offers significant advantages in this context. In particular, the explicit articulation of the inferential steps required by the Bayesian framework enhances the transparency of the reasoning process and makes it fully amenable to scrutiny within the adversarial process. The fact that the construction of the model necessarily involves discretionary choices does not imply arbitrariness or reliance on mere intuition, since those choices remain constrained by standards of intersubjective rationality[167]. Nor should it be forgotten that these are assessments which, to a considerable extent, the judge would in any event be required to make. The distinctive value of the model lies precisely in the fact that such assessments are formalized and thereby rendered transparent and verifiable. In this respect, the system is further strengthened by the duty to provide reasons for the use of these tools in the judgment, which enables the parties to scrutinize the assumptions underlying the model and, where appropriate, to challenge on appeal any illogical, unsupported, or inconsistent inferential choices.

These considerations bring into focus two particularly significant points. First, they demonstrate the compatibility of Bayesian methods with the requirements of transparency and meaningful human control that, as discussed above, constitute essential safeguards for the legitimate use of algorithms in criminal proceedings. Second, they reveal that the Italian legal system already possesses mechanisms capable of preventing a merely intuitive or uncritical use of such methodologies, including in their computational form, precisely because the underlying assessments remain subject to rigorous rational scrutiny by all the actors involved in the proceedings.

---

[165] In this regard, cf., MACKOR, *Bayesian Modelling*, cit., 369 ff.

[166] GENNARI, *I giudici sono pronti per l'Activity Level*, cit., rightly emphasizes this point.

[167] This point is rightly noted by GENNARI, *I giudici sono pronti per l'Activity Level?*, cit.

3.2. *The Crime Level: Where Should the "Pillars of Hercules" of Trial by Probabilities Be Placed?* The most delicate question that remains to be addressed concerns the use of Bayesian methods in relation to propositions situated at the crime level. Here, the possible applications of the Bayesian framework are far from homogeneous. Rather, they vary considerably depending on several factors.

The first distinction concerns the actors involved: at least in principle, Bayesian reasoning may be employed by experts, by the parties, or even directly by the trial judge[168].

A second distinction relates to the object and modality of its use. As noted above, the theorem may operate merely as a logical framework for structuring evidentiary reasoning through "analogical" calculations performed by the human decision-maker, or it may be implemented through computational tools which, by means of Bayesian networks, assist in the analysis and combination of evidence.

Finally, applications differ in terms of their intensity. The Bayesian framework may be used to assess a single item of evidence or a limited set of proofs, but it may also be extended, in more ambitious scenarios, to the overall structuring of evidentiary reasoning as a whole[169].

In its strongest form, the Bayesian approach could lead to the transformation of standards of proof into quantitative thresholds of rational belief, determined through normative choices (for example, 99% for conviction), beyond which the decision-maker would be entitled to regard the relevant hypothesis as proven[170]. From this perspective, Bayes' theorem would operate as a formal instrument for determining whether the prescribed threshold has been reached. The significance of such an approach is readily apparent. If feasible, it would offer a potential solution to one of the most debated problems in the theory of standards of proof, namely the difficulty of expressing different levels of proof through linguistic formulas and the interpretive indeterminacy

[168] See MACKOR, *Bayesian Modelling*, cit., 376 ff. Cf., instead, SALLAVACI, *Algorithms on Trial*, cit., 9, who remembers that «it is a well-established principle of forensic evidence evaluation approach that offence level propositions are strictly for the fact finder to consider».

[169] Cf. again the recent analysis by MACKOR, *Bayesian Modelling*, cit., 376 ff.

[170] For an attempt to estimate the threshold for criminal conviction at a probabilistic level, see CHERUBINI, *Guida al ragionamento probatorio*, cit., 391 ff., as well as GARBOLINO, *Probabilità e logica della prova*, cit., 461 ff.

that inevitably follows from them[171]. Quantitative formalization would, in principle, make it possible to verify *ex post*, in a far more precise manner, whether the required evidentiary threshold had actually been attained.

However, the more Bayesian methodology is extended to strong and highly automated forms of application at the crime level, the more it gives rise to a series of complex issues that have long been emphasized both in the literature[172] and in case law[173].

Among the most frequently advanced objections are, first, concerns relating to the intelligibility and practical feasibility of mathematical approaches, particularly in systems involving lay decision-makers such as juries. Secondly, critics emphasize the need for criminal adjudication to remain consistent with widely shared conceptions of procedural fairness and moral legitimacy, a requirement that is difficult to reconcile with any attempt to reduce the trial to a form of "trial by mathematics"[174].

A further source of concern lies in the operational difficulties that arise whenever no sufficiently robust statistical basis exists for estimating the relevant likelihoods. In such circumstances, the risk of the so-called "quantification paradox" becomes particularly evident. The problem consists in attempting to express, in numerical terms, degrees of belief relating to individual facts, thereby conferring an appearance of precision upon assessments that remain inherently uncertain and, to a significant extent, qualitative in nature[175].

---

[171] In this regard, still relevant is WIGMORE, A *Treatise on the Anglo-American System of Evidence in Trials at Common Law*, vol. IX, New York, 1940, 325, who – starting from the premise according to which «the truth is that no one has yet invented or discovered a mode of measurement for the intensity of human belief» - arrived at the conclusion according to which «there can be yet no successful method communicating intelligibly to a jury a sound method of self-analysis of one's belief».

[172] From a comparative point of view, fundamental is the work by TRIBE, *Trial by mathematics: precision and ritual in the legal process*, in *Harvard Law Review*, 1971, 84(6), 1329 ff. See also, *ex multis*, ALLEN-PARDO, *The Problematic Value of Mathematical Models of Evidence*, in *Journal of Legal Studies*, 2007, 1, 107 ff.; COHEN, *The Probable and the Provable*, Oxford, 1977; FERRER BELTRÁN, *La valutazione razionale della prova*, cit., 114 ff.; HAACK, *Evidence Matters, Science, Proof, and Truth in the Law*, Cambridge, 2014, 56 ff.; LAUDAN, *Truth, Error, and Criminal Law*, Cambridge, 2006, 100 ff.; TARUFFO, *La prova dei fatti giuridici*, Milan, 1992, 199 ff.

[173] See, for instance, Supreme Court of the Connecticut *State v. Skipper*, 637 A.2d 1101, as well as the Court of Appeal for England and Wales, *R v Adams* (No 1) [1996] 2 Cr App R 467; *R v Adams (No.2)* [1998] 1 Cr.App.R 377 e *R v. T* [2010] EWCA 2439.

[174] Cf. TRIBE, *Trial by mathematics*, cit., 1372.

[175] See TRIBE, *Trial by mathematics*, cit., 1361 f.

The most radical objection, however, concerns the justification of prior probabilities within Bayesian reasoning. It has been argued that attributing a positive prior probability to the hypothesis of guilt is difficult to reconcile with the presumption of innocence[176]. This objection gives rise to structural difficulties for strong applications of Bayes' theorem at the crime level: if the prior probability of guilt were set at zero, the calculation could never yield a positive posterior probability, since any value multiplied by zero necessarily remains zero. In such circumstances, the theorem would be deprived of any practical utility[177].

It is against this background that a number of alternative theories of evidentiary reasoning have been developed, emphasizing qualitative and argumentative modes of justification over formal probabilistic calculation. Among these, some jurisdictions have increasingly embraced approaches based on the comparison of competing factual scenarios (the so-called scenario theory)[178]. In others, greater influence has been exerted by the notion of "Baconian" or "logical" probability, understood not as an expression of the decision-maker's subjective degree of belief in the truth of a proposition, but rather as a measure of the degree of confirmation objectively supported by the available evidence[179].

As is well known, approaches grounded in the notion of "logical probability" have gained broad acceptance not only in legal scholarship[180] but also in Italian case law[181]. Indeed, while Bayesian reasoning has found applications that are

[176] For a resume of the critics, see FERRER BELTRÁN, *La valutazione razionale della prova*, cit., 119 f.

[177] Cf., in very explicit terms, *State v. Skipper*, 637 A.2d 1101, in which the court strikingly stated that Bayes' theorem could operate only if the presumption of innocence were disregarded.

[178] Concerning this approach, see MACKOR-VAN KOPPEN, *The scenario theory about evidence in criminal law*, in *Philosophical Foundations of Evidence Law*, ed. by Dahlman-Stein-Tuzet, Oxford, 2021, 215 ff.

[179] The leading work of this line of interpretation is COHEN, *The Probable and the Provable*. cit., *passim*. Among the supporters of this interpretation, see FERRER BELTRÁN, *La valutazione razionale della prova*, cit., 119 f.; STELLA, *Il giudice corpuscolariano. La cultura delle prove*, Milan, 2005, 225 f.; TARUFFO, *La prova dei fatti giuridici*, cit., 199 ff.

[180] See, for instance, CAPRIOLI, *L'accertamento della responsabilità penale "oltre ogni ragionevole dubbio"*, in *Riv. it. dir. proc. pen.*, 2009, 1, 66 ff. and MAZZA, *Il ragionevole dubbio nella teoria della decisione*, cit., 361 ff.

[181] Cf., *inter alia*, Cass. civ., Sec. III, 11 June 2024, No. 16199, in *One Legale*, as well as Cass. civ., Sec. III, 7 April 2021, No. 32860, *ibid.* For a critical perspective on the judicial misuse of this concept, see, however, the incisive remarks by BONIOLO-GENNARI, *Note su giurisprudenza e probabilità: fra leggi di natura e causalità*, in *Sist. Pen.*, 2021, 10, 103 f., as well as by STELLA, *Il giudice corpuscolariano*, cit., 226.

far from negligible in civil litigation (particularly in the field of paternity determination)[182], a different conception of evidentiary reasoning has become dominant in criminal proceedings following the landmark *Franzese* judgment[183]. According to this approach, standards of proof are not conceived as quantitative thresholds expressed in numerical terms. Rather, they are understood as ordinal criteria, constructed through a process of "inductive elimination"[184], whereby competing hypotheses are progressively tested and excluded until any reconstruction reasonably compatible with the defendant's innocence has been ruled out[185].

At the same time, the dominant trend in Italian criminal case law has gone beyond merely endorsing the model of logical probability. In the few decisions that have directly addressed the issue, the Supreme Court has expressed clear reservations about the use of Bayes' theorem in criminal adjudication. A significant example is provided by a 2018 judgment of the Court of Cassation, which rejected a defense argument based on Bayesian reasoning and intended to demonstrate the logical implausibility of the prosecution's reconstruction of the facts. In this regard, the Court held more specifically that the theorem is designed to assess the probability of repeatable events whose frequency is known, whereas it is ill-suited to the reconstruction of «human conduct at a given historical moment», since human behavior may also follow «deviant» or «irrational» patterns that escape repetitive statistical regularities[186].

However, as has been convincingly observed in the legal literature, this line of reasoning appears to rest on a significant conceptual misunderstanding[187]. In the judgment discussed above, the Supreme Court seems to have conflated frequentist probability – grounded in the statistical regularity and repeatability of events – with Bayesian probability, which is instead specifically designed to

---

[182] See Cass. civ., 11 December 1980, n. 6400, in *One Legale*. See also Trib. Minorenni Catania, 24 October 1991, *ivi*; Trib. Minorenni Venezia, 27 November 1986, *ivi*.

[183] Cf. Cass., Joint Sections, 10 July 2002, Franzese, in *Riv. it. dir. proc. pen.*, 1133 ff., which authoritatively clarified that, whereas "statistical probability" concerns the empirical frequency with which a given event occurs within a class of events, "logical probability", which is relevant to criminal adjudication, requires an additional assessment of whether that statistical generalisation can reliably be applied to the individual case in light of the available evidence.

[184] In this regard, also for further doctrinal references, see DELLA TORRE, *Taking the Evolution*, cit., 195.

[185] See MAZZA, *Il ragionevole dubbio nella teoria della decisione*, cit., 366 ff.

[186] Cf. Cass., Sec. II, 7 May 2019, n. 19138, in *One Legale*.

[187] See VALENTINI, *La "scoperta" dei* bias *cognitivi*, cit., 9.

assess singular events through the updating of prior beliefs in light of the available evidence[188].

The very uncertainty displayed in distinguishing between different conceptions of probability – even at the level of the highest court – is sufficient to cast doubt on the view that the application of Bayes' theorem at the crime level should be regarded as categorically precluded. Indeed, the theorem is by no means inherently incompatible with the assessment of individual historical facts. On the contrary, it may provide a valuable tool for avoiding misunderstandings in probabilistic reasoning, including those that have occasionally surfaced in the jurisprudence of the Court of Cassation itself.

This conclusion appears all the more convincing when one considers that the Bayesian approach has not remained static, but has instead given rise to a rich and sophisticated body of scholarship aimed at addressing the principal objections traditionally raised against it[189].

In this regard, it should first be recalled that concerns relating to the complexity and intelligibility of Bayesian reasoning have been addressed, at least in part, through the development of computational tools such as Bayesian networks. These tools facilitate the performance of probabilistic calculations through relatively user-friendly software while also enabling complex inferential scenarios to be represented graphically in a more transparent and comprehensible manner[190].

Nor has the debate remained static regarding the crucial issue of prior probabilities. In addition to the development of models designed to reconcile their use with the presumption of innocence[191], a variety of theoretical solutions have been proposed to provide a rational basis for their determination[192].

It should be noted that, as part of the literature has observed, the growing datafication of contemporary society opens up new possibilities for refining such

---

[188] Even among the most ardent critics of the Bayesian approach, this distinction is clearly acknowledged; see, for example, FERRER BELTRÁN, *La valutazione razionale della prova*, cit., 95 ff.

[189] For a reply to the criticisms to the bayesian method, cf. GARBOLINO, *Probabilità e logica della prova*, cit., 134 ff.

[190] See GARBOLINO, *Probabilità e logica della prova*, cit., 137.

[191] Cf., in particular, CHERUBINI, *Guida al ragionamento probatorio*, cit., 357 ff.; FRIEDMAN, *A Presumption of Innocence, Not of Even Odds*, in *Stanford Law Review*, 2000, 4, 873 ff.; SALLAVACI, *Algorithms on Trial*, cit., 9 ff.

[192] In this regard, for a resume, cf. DAHLMAN-KOLFLAATH, *The Problem of the Prior in Criminal Trials*, in *Philosophical Foundations*, cit., 287 ff., as well as FENTON-LAGNADO-DAHLMAN-NEIL, *The Opportunity Prior - A Proof-Based Prior for Criminal Cases*, in *Law, Probability and Risk*, 52(4), 2019, 237 ff.

assessments by making available increasingly rich informational bases from which to derive probabilistic estimates, particularly with regard to likelihoods and contextual factors[193]. It is precisely in light of these developments that some commentators have envisioned the emergence of legal systems increasingly shaped by forms of "algorithmic legal proof", in which artificial intelligence and Bayesian network models could come to dominate evidentiary reasoning, thereby posing a serious challenge to the traditional freedom of judicial evaluation[194].

However, the risk of such scenarios materializing – which, at present, remains largely prospective – is mitigated by the safeguards already discussed and now codified in the European AI Regulation. It must indeed be acknowledged that the algorithmic implementation of Bayesian methodology through software based on Bayesian networks appears to fall within the category of AI systems intended to be used by, or on behalf of, judicial authorities to assist in the establishment and interpretation of facts, pursuant to Annex III, point 8(a), of the Regulation. As such, these applications qualify as "high-risk" AI systems[195]. However, this classification does not imply any absolute prohibition of their use. Rather, it means that any deployment of such tools must comply with the stringent requirements applicable to high-risk systems, particularly those concerning data quality and governance, transparency of the methodologies employed, effective human oversight, and the accuracy, robustness, and reliability of the system. These are, moreover, requirements that a careful and methodologically rigorous use of Bayesian models appears, at least in principle, capable of satisfying.

The foregoing considerations suggest that, although "strong" applications of Bayesian methodology – capable of transforming standards of proof into rigid quantitative thresholds – remain difficult to envisage at present, an *a priori* rejection of such tools may prove equally problematic. The danger is that such a stance may reflect, on the one hand, an incomplete understanding of the model and, on the other, an implicit preference for decisional frameworks that preserve particularly broad margins of evaluative discretion. This is all the truer if one considers a point that is often overlooked: Bayes' theorem does not ordinarily require the decision-maker to perform additional inferential

[193] The reference goes to the thesis adopted by TUZET, *La prova ragionata*, cit., 244 f.
[194] Cf., again, TUZET, *La prova ragionata*, cit., 244 f.
[195] With regard to this prevision, see DELLA TORRE, *Cripto-attività e giustizia penale*, cit., 398.

operations. Rather, it requires those inferential steps that are already inherent in rational evidentiary reasoning to be made explicit, structured, and open to scrutiny.

In light of the foregoing considerations, an intermediate position – already advanced in the literature – appears particularly promising. By moving beyond the sterile dichotomy between explanatory approaches and Bayesian models[196], it advocates their integration into a unified framework in which probabilistic tools and non-numerical criteria operate in a complementary fashion to enhance the control and rationality of evidentiary reasoning[197].

In this regard, it should be emphasized that such a proposal offers a significant methodological advantage. By integrating Bayesian methodology with qualitative criteria – such as coherence, completeness, economy, and the explanatory and predictive capacity of competing hypotheses[198] – it becomes possible to combine the strengths of both approaches within a unified evidentiary framework. Indeed, the integration of the two approaches allows them to perform distinct but complementary functions. Qualitative criteria would play a primary role in the generation, refinement, and selection of the competing hypotheses. Bayesian methodology, in turn, supports their comparative evaluation by structuring and making the inferential process more transparent, facilitating the organisation of complex inferential relationships – particularly through visual representations[199] –and reducing the risk of inferential errors and logical fallacies. The overall result is a more epistemically rigorous meth-

[196] In this regard, see TILLERS, *Trial by mathematics – reconsidered*, in *Law, Probability and Risk*, 2011, 10(3), 167 ff.

[197] Cf., for instance, MACKOR, *Bayesian Modelling*, cit., 377, who, together with C. Dahlman and D. Lagnado, is developing the research project *NWO "Prevention Miscarriages of Justice" (PMJ)* (no. 406.21.RB.004), entirely devoted to these themes. See also JELLEMA, *(Im)probable stories: combining Bayesian and explanation-based accounts of rational criminal proof*, PhD Thesis, University of Groningen, 2023. In the Italian literature, a similar position is advanced by MAESTRI, *Verità giudiziaria e intelligenza artificiale. Dalla prova baconiana alla meta-prova algoritmica*, in *Arch. pen. web*, 2025, 3, 5 f. and 28.

[198] It is no coincidence that these criteria are also emphasized by scholars such as FERRER BELTRÁN, *Prova senza convincimento*, cit., 198 ff., who advocates a qualitative and linguistic conception of standards of proof.

[199] Indeed, one of the principal advantages of Bayesian networks lies precisely in their ability to complement the argumentative paths developed by the parties, whether in written submissions or oral argument, with a visual representation of competing hypotheses. In this way, the comparative assessment of those hypotheses becomes more immediate and transparent, while reducing the risk that relevant alternatives may be overlooked or disregarded altogether.

od of assessing evidence, while avoiding the risk of reducing judicial decision-making to a purely numerical exercise.
Moreover, it should be emphasized that, provided it is not pushed to its more radical forms, the value of the Bayesian approach lies not primarily in the numerical results it produces – which must in any event remain grounded in assumptions capable of intersubjective verification – but rather in its ability to make explicit, within a transparent and structured model of reasoning, the relationships between the evidence and the competing hypotheses[200]. Its principal contribution lies precisely in this capacity: by rendering inferential links visible and open to scrutiny, Bayesian methodology facilitates a more controlled assessment of hypotheses, both individually and comparatively, while reducing the risk that crucial stages of evidentiary reasoning remain implicit, insufficiently justified, or shielded from critical examination.
The potential of this approach also finds significant empirical support in one of the few Italian decisions on the merits in which Bayes' theorem was explicitly applied at the crime level[201]. In that case, the Tribunal of Milan relied on Bayesian reasoning not merely to avoid overestimating the probative value of scientific evidence, but also to demonstrate that the evidentiary record as a whole – comprising both technical-scientific findings and qualitative circumstantial elements – was insufficient to exclude a plausible alternative hypothesis consistent with the defendant's innocence. As a result, the evidentiary threshold required by Article 533 CCP could not be regarded as satisfied. Particularly noteworthy is the role assigned to numerical values in that decision. They were not treated as the expression of an exhaustive quantification of evidentiary assessment, but rather as argumentative and explanatory tools embedded within a broader reasoning process and subject to rational scrutiny. In this respect, the decision appears fully consistent with the idea that «probability is not about numbers; it is about reasoning»[202].
This does not, of course, mean that all the difficulties associated with the Bayesian approach have been resolved. As acknowledged even by some of the leading proponents of the intermediate perspective outlined above, sever-

[200] Cf. BONIOLO-GENNARI, *Note su giurisprudenza e probabilità*, cit., 104, who observe that the objective should not be to open the door to "automated" decision-making, but rather to provide a verifiable structure for judgments whose inferential foundations often remain implicit.
[201] The reference goes to Tribunal of Milan, sez. G.I.P., 18 June 2015, giud. Gennari, cit.
[202] Cf. L. LUPÁRIA DONATI, Trials by probabilities, cit., 98.

al important issues remain open[203]. These range from the problem of reliably determining likelihoods and prior probabilities to the question of who is best placed to construct and manage Bayesian networks, particularly in complex cases. At present, for example, it appears difficult to maintain that such tasks could be performed directly by the judge, making the involvement of experts seem indispensable. At the same time, an active role for the parties cannot be ruled out, both in the construction of the models and in their adversarial examination and critique. It is equally evident that, from a longer-term perspective, the meaningful integration of such tools into judicial practice would require significant investment in training and capacity-building. This would entail fostering a more widespread understanding of probabilistic reasoning and its implications among legal professionals, including through the introduction of specialized educational programs already tested in other European jurisdictions[204].

Nevertheless, these difficulties do not diminish the force of the invitation, advanced by leading scholars more than a decade ago, to overcome the excessive distrust and prejudice that still surround instruments for the formalization of evidentiary reasoning[205]. Indeed, if the fundamental aspiration of a criminal justice system committed to fairness is to improve the epistemic quality of its decisions, thereby reducing the risk of miscarriages of justice and facilitating the correction of those that occur, then a controlled integration of Bayesian tools into evidentiary assessment should be regarded as a valuable opportunity rather than a threat. In this sense, properly understood and carefully employed, these instruments have the potential to make evidentiary reasoning more rigorous, structured, and amenable to scrutiny, thereby strengthening the protection afforded to the innocent[206].

---

[203] See MACKOR, *Bayesian Modelling*, cit., 382 f.

[204] This is the case, for example, in the Netherlands, where specialised training programmes for members of the judiciary are currently being developed as part of the above-mentioned *Prevention of Miscarriages of Justice* project. More generally, on the importance of judicial training in this field, as well as on the potential benefits of introducing judicial assistants with expertise in probabilistic reasoning, see MACKOR, *Probability clerks and Probability Judges. Or How to Prevent Probabilistic Fallacies in Court*, in *Legal Argumentation: Reasoned dissensus and common ground*, ed by E. Feteris et. al., Den Haag, 2024, 114 f.

[205] See LUPÁRIA DONATI, Trials by probabilities, cit., 100 f.

[206] On this point, see also VALENTINI, *La "scoperta" dei* bias *cognitivi*, cit., 10.