ATTUALITÀ

PIETRO MIRTO

La dematerializzazione della sanzione e della prova: l'illusione del vero algoritmico e la tenuta del «ragionevole dubbio» nell'era dei *deepfake*

Il contributo analizza l'impatto dirompente delle tecnologie di intelligenza artificiale generativa con specifico riguardo alla fenomenologia dei *deepfake* sui nodi strutturali del diritto penale, sia sul versante sostanziale sia su quello processuale. Sotto il profilo dogmatico, l'indagine vaglia la tenuta e l'adeguatezza delle fattispecie incriminatrici tradizionali, nonché dei recenti interventi riformatori, nella tutela dei beni giuridici della personalità e della pubblica fede, evidenziandone i deficit in termini di tassatività e determinatezza strutturale. Sul versante strettamente processualistico, la riflessione si focalizza sulla crisi epistemologica della prova digitale: l'avvento di alterazioni *cross*-mediali ormai indistinguibili dal dato reale mina la genuinità del compendio probatorio e impone una ridefinizione dei confini operativi della regola di giudizio dell'«oltre ogni ragionevole dubbio» (*ex* art. 533 c.p.p.). In chiave *de jure condendo*, l'autore prospetta una necessaria revisione dei criteri di ammissibilità, sbarramento e autenticazione della *e-evidence*, auspicando una sinergica integrazione tra categorie penalistiche e protocolli crittografici di tracciabilità e immutabilità del dato alla fonte.

The Dematerialisation of Sanctions and Evidence: The illusion of Algorithmic Truth and the Resilience of «Reasonable Doubt» in the era of Deepfakes

The paper analyses the disruptive impact of generative artificial intelligence technologies, with specific regard to the phenomenon of deepfakes, on the structural nodes of criminal law, both from a substantive and a procedural perspective. From a dogmatic standpoint, the investigation examines the resilience and adequacy of traditional criminal offences, as well as recent reform interventions, in protecting the legal interests of personality rights and public trust, highlighting their shortcomings in terms of strict legality and structural certainty. On the strictly procedural side, the reflection focuses on the epistemological crisis of digital evidence: the advent of cross-media alterations that are now indistinguishable from real data undermines the authenticity of the body of evidence, it forces a redefinition of the operational boundaries of the "beyond any reasonable doubt" standard of judgment (art. 533 of the Italian Code of Criminal Procedure). From a de jure condendo perspective, the author proposes a necessary revision of the criteria for the admissibility, filtering, and authentication of e-evidence, advocating for a synergistic integration between criminal law categories and cryptographic protocols for traceability and immutability of data at the source.

1. *Il collasso epistemologico del dato empirico nell'orizzonte della* post *verità processuale.* L'epistemologia del processo penale di matrice illuministicaa si radica sull'indefettibile postulato della corrispondenza empirica tra il fatto sto-

rico oggetto di contestazione e la sua rappresentazione all'interno del perimetro dibattimentale.[1] La scommessa del rito accusatorio, storicamente orientata alla ricerca di una verità processuale che si approssimi quanto più possibile alla verità materiale attraverso il metodo del contraddittorio, subisce oggi una torsione radicale indotta dall'avvento della cosiddetta intelligenza artificiale generativa.[2] La fenomenologia dei *deepfake* intesi quali sintesi *cross*-mediali di flussi audio-visivi operati mediante reti neurali generative avversarie (GAN) non si configura come una mera evoluzione quantitativa delle tradizionali tecniche di contraffazione documentale.[3] Al contrario, si assiste a una mutazione ontologica della prova: la tecnologia cessa di essere un mero strumento di alterazione del supporto analogico o digitale preesistente per divenire una vera e propria fabbrica di realtà simulate, dotate di un'intrinseca forza persuasiva biologicamente e visivamente indistinguibile dal vero storico.[4] Il pericolo latente, che la dottrina processuale penalistica più avveduta inizia a delineare, consiste nel passaggio da un sistema di libero convincimento del giudice ancorato a criteri logico-razionali ed empirici, a un regime di scetticismo radicale o, di contro, di fede tecnologica acritica.[5] Se ogni traccia audiovisiva, storicamente considerata dotata di una formidabile carica di immediatezza rappresentativa, diviene manipolabile alla radice in modo impercettibile per i sensi umani e per i comuni strumenti diagnostici forensi, l'intero statuto della

[1] Sul concetto di verità processuale e sui suoi limiti epistemologici nella transizione dal modello inquisitorio a quello accusatorio, sia consentito il rinvio fondamentale a FERRUA, *Il processo penale: profilo teorico e sistematico, Torino,* GIAPPICHELLI, *2022, p. 45 ss.;* nonché, in una prospettiva filosofico-giuridica, FERRAJOLI, *Diritto e ragione. Teoria del garantismo penale, Roma-Bari, Laterza, 1989, p. 412.*

[2] UBERTIS, *Il sistema della prova penale, Milano, Giuffrè, 2022, p. 115 ss.;* nonché *Epistemologia, intelligenza artificiale e decisione giudiziale, in Rivista italiana di diritto e proc. pen., 2024, fasc. 2, p. 445 ss.,* il quale analizza i limiti della razionalità del giudizio logico a fronte della scomposizione algoritmica del dato empirico.

[3] Per il saggio seminale sul funzionamento delle reti GAN *(Generative Adversarial Networks), vedi* GOODFELLOW et al., *Generative Adversarial Networks,* in *Advances* in *Neural Information Processing Systems, 27, 2014, p. 2672 ss.* Sulla dematerializzazione del supporto nell'era della riproducibilità algoritmica, LUPÀRIA, *L'evidenza digitale nel processo penale italiano, Milano, Giuffrè, 2021.*

[4] Sul forte impatto cognitivo delle prove visive e sul rischio di distorsione del convincimento del decisore, CARLIZZI, *Forma informatica e prova giudiziale, Milano, Giuffrè, 2018, p. 95 ss.*

[5] Sul rischio latente che il giudicante abdichi al proprio ruolo critico affidandosi passivamente alla regolarità formale dello strumento tecnologico PANZAVOLTA, *L'era della giustizia digitale: decisioni algoritmiche e garanzie processuali, in Cass. pen, 2023, p. 1120;* nonché CAMALDO, *La formazione della prova informatica tra regole tecniche e libero convincimento, Milano, Giuffrè, 2024, p. 78.*

e-evidence entra in crisi di rigetto.[6] Non si tratta soltanto di ridefinire le regole di ammissibilità o di valutazione della prova documentale ex art. 234 c.p.p., bensì di scongiurare il collasso della funzione cognitiva della giurisdizione.[7] Il giudice rischia di trovarsi sospeso in un limbo interpretativo descritto da una dottrina anglosassone come *infocalypse*: uno scenario in cui l'incertezza sulla genuinità del dato informatico mina l'affidabilità dell'intero materiale probatorio d'accusa, provocando un effetto paralizzante sull'accertamento giudiziale o, peggio, aprendo la strada a derive di puro arbitrio decisionale.[8] È in questo preciso iato che si consuma, peraltro, la dematerializzazione della sanzione: l'effetto dirompente e irreversibile della diffusione del falso algoritmico opera un'anticipazione della dimensione punitiva nella sfera mediatico-sociale, slegandola dalla materialità dell'esecuzione penale (art. 27 Cost.) e trasformando il processo stesso – inficiato dal virus dell'incertezza in una sanzione cognitiva autonoma e non emendabile. Dinanzi alla mutazione ontologica impressa dai *deepfake*, la prima tentazione dell'interprete orientato al positivismo legislativo è quella di rifugiarsi nell'alveo delle riforme strutturali che hanno disciplinato la *e-evidence*.[9] Il riferimento primario corre alla L. 18 marzo 2008, n. 48 (di ratifica della Convenzione di Budapest sul crimine informatico) e all'introduzione dell'art. 234-bis c.p.p., norma nata per regolamentare l'acquisizione di documenti informatici conservati all'estero o canali di comunicazione digitale. Tuttavia, questo impianto normativo svela oggi una drammatica inadeguatezza. La L. 48/2008 si preoccupa di garantire la ripetibilità e la non alterabilità del dato informatico dal momento del suo reperimento in poi. L'accento è posto sulle buone prassi forensi di copia speculare (*bit-stream image*) e sulla cristallizzazione del valore di *hash*.[10] Si delinea così il pa-

---

[6] In ordine alla crisi d'identità della prova informatica, si veda l'ampia disamina di SIGNORATO, *L'acquisizione della prova digitale nel processo penale, Padova, CEDAM, 2020, p. 55 ss.*

[7] Sul punto, CAIANIELLO, *Premesse per una teoria della prova nel processo penale europeo,* in *Rivista italiana di diritto e procedura penale, 2022, fasc. 3, p. 911.*

[8] Il termine infocalypse è mutuato dalla sociologia dei media e applicato all'epistemologia giudiziaria da CHESNEY-
CITRON, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security,* in *California Law Review, vol. 107, 2019, p. 1753.*

[9] Per una ricostruzione dell'evoluzione normativa in materia di e-evidence, CINGARI, *Informatica forense e processo penale, Torino,* GIAPPICHELLI, *2021.*

[10] Sul rigore metodologico della *computer forensics* e sui limiti della cristallizzazione del dato tramite *hashing*, si veda l'ampia disamina di CURTOTTI–SARZANA DI SIDERCI, *Trattato di digital forensics. Tecnologie investigative, prove digitali e sicurezza informatica, Padova, CEDAM, 2017, p. 245 ss.*

radosso dell'immutabilità del falso: se un flusso video immesso nel circuito d'indagine è l'esito di una manipolazione operata *ab origine* tramite reti *GAN,* la polizia giudiziaria che applicherà ortodossamente i protocolli di *computer forensics* non farà altro che congelare, repertare e garantire l'immutabilità di un perfetto falso storico. L'integrità digitale (l'assenza di modifiche successive al sequestro) viene contrabbandata con la genuinità storica (la corrispondenza al vero dell'evento rappresentato).[11] L'art. 234-bis c.p.p., limitandosi a disciplinare l'acquisizione della prova digitale, non offre alcun criterio di valutazione epistemologica del suo contenuto intrinseco. Si assiste a una pericolosa regressione del metodo giudiziale: il giudice, abbagliato dalla conformità formale delle operazioni peritabili, rischia di abdicare al proprio ruolo di *peritus peritorum* per scivolare nell'*automation bias*, assumendo acriticamente che un file formalmente integro sia per ciò solo storicamente vero. La Suprema Ccorte di cassazione ha affrontato a più riprese il valore probatorio delle riproduzioni informatiche, ma ancorandosi a categorie ermeneutiche strutturate sull'antico paradigma analogico. Sotto un primo profilo, il consolidato orientamento di legittimità assimila le registrazioni fonografiche o videografiche ai documenti *ex* art. 234 c.p.p., stabilendo che la loro efficacia probatoria possa essere incrinata solo da una contestazione specifica e circostanziata della parte interessata, e non già da generiche doglianze. Sotto un secondo profilo, in tema di reati che traggono il proprio innesco probatorio principale da videoriprese si pensi ai reati di pornografia minorile o ai procedimenti di criminalità organizzata, la giurisprudenza ha costantemente affermato che, in assenza di macroscopici elementi di sospetto o di una formale e dettagliata allegazione difensiva circa la manipolazione, il filmato reca in sé una piena valenza probatoria, potendo il giudice fondare su di esso il proprio convincimento senza necessità di previo accertamento peritale. Questo spartito esegetico, se applicato ai *deepfake*, espone il sistema a un rischio sistemico di errore giudiziario. Le reti neurali avversarie operano per l'appunto eliminando gli artefatti *pixel* per *pixel*, rendendo la falsificazione invisibile ai sensi umani. Pretendere che la difesa individui "macroscopici elementi di sospetto" visivi significa esigere una facoltà divinatoria o, sul piano squisitamente processuale, configurare una insostenibile *probatio diabolica* in aperto contrasto con il diritto di difesa. Per comprendere la portata destabilizzante del fenomeno, oc-

[11] CAMON, *La prova informatica, in Sistemi penali, 2020, n. 11, p. 5 ss.*

corre calare le coordinate teoriche sinora tracciate all'interno della dinamica processuale, isolando tre direttrici fenomenologiche paradigmatiche. La prima direttrice concerne l'ipotesi della fabbricazione algoritmica del materiale d'accusa. Si pensi allo scenario in cui un collaboratore di giustizia consegni all'organo inquirente un *file* video, asseritamente estratto da una piattaforma di messaggistica cifrata, che ritragga un esponente politico nel momento di ricevere una dazione indebita di denaro da un esponente apicale della criminalità organizzata. Là dove l'estrazione forense si riveli formalmente impeccabile ai sensi della L. n. 48/2008, l'impatto cognitivo sul giudice cautelare risulterà devastante, quand'anche il filmato sia, in realtà, un *deepfake* generato mediante la sovrapposizione digitale del volto del soggetto indagato sul corpo di un terzo. In simile evenienza, l'asimmetria economico-tecnologica necessaria a sostenere i costi di una consulenza tecnica idonea a rilevare le microasincronie biometriche finisce per traslare illegittimamente l'onere della certezza in capo all'imputato. Specularmente, una seconda linea di crisi si manifesta nell'alveo delle strategie difensive volte alla distruzione dell'alibi. È il caso dell'imputato per un delitto di sangue che produca in dibattimento un *file* video, apparentemente estratto da un circuito di videosorveglianza privata, che lo ritragga all'interno di un esercizio commerciale a rilevante distanza dal *locus commissi delicti* al momento della consumazione del reato. Dinanzi al sospetto di un falso algoritmico, in assenza di strumenti diagnostici di immediata e certa computazione in capo alla pubblica accusa, l'operatività del canone dell'*in dubio pro reo* rischia di imporre un'assoluzione obbligata, paralizzando la pretesa punitiva dello Stato sulla scorta di una realtà simulata. Infine, non meno insidioso appare il fenomeno che la dottrina sociologico-giuridica definisce quale *liar's dividend* (il dividendo del bugiardo), ovverosia lo sfruttamento dello scetticismo radicale indotto dall'avvento dell'intelligenza artificiale generativa. Si pensi all'imputato del delitto di violenza sessuale attinto da un filmato registrato dallo *smartphone* della vittima durante l'azione delittuosa, il quale si limiti a eccepire formalmente la natura artificiale del supporto. In quest'ultimo scacchiere emerge il risvolto simmetrico dell'*infocalypse*: l'erosione sistemica della fiducia nella tenuta epistemica della prova scientifica consente al colpevole di ammantare di dubbio tracce probatorie totalmente genuine, frustrando l'accertamento della verità materiale. Per disinnescare la crisi e offrire una soluzione interpretativa che tenga uniti il principio di colpevolezza oltre ogni ragionevole dubbio (art. 533 c.p.p.) e la tutela

dell'affidamento nella giurisdizione, è necessario tracciare uno statuto ermeneutico bifasico, fondato sul dialogo tra gli artt. 24, 27, co. 2 e 111 della Carta costituzionale. In primo luogo, occorre perimetrare un onere qualificato di allegazione (e non di prova) a carico della difesa. Per evitare l'effetto paralizzante del *liar's dividend*, la mera evocazione astratta dello "spauracchio del *deepfake*" non può essere ritenuta sufficiente a incrinare il compendio probatorio. La difesa non deve dimostrare scientificamente la falsità del supporto operazione che configurerebbe un'incostituzionale inversione dell'onere della prova, ma deve offrire al giudice elementi di specifica plausibilità della manipolazione, quali l'oscurità dei canali di provenienza del *file*, l'assenza di testimoni o la stridente contraddittorietà logica con altri atti del procedimento. Siffatto onere non viola il principio di presunzione di innocenza: una volta assolto l'onere di allegazione difensiva mediante la prospettazione di un dubbio plausibile e contestualizzato, l'onere della prova contraria si ridistribuisce interamente in capo al Pubblico ministero, sul quale graverà l'obbligo di dimostrare, oltre ogni ragionevole dubbio, la genuinità della traccia digitale e l'integrità della sua catena di custodia.

2. La faglia sostanziale: la tutela dei beni giuridici individuali e collettivi alla prova delle fattispecie di nuovo conio. Lo sfilacciamento epistemologico descritto sul piano processuale riverbera i suoi effetti simmetrici sul versante del diritto penale sostanziale, ponendo in crisi il principio di frammentarietà e, soprattutto, l'obbligo costituzionale di tassatività e determinatezza del precetto penale (ex art. 25, co. 2, Cost.).[12] Dinanzi alla proliferazione di condotte lesive realizzate mediante la creazione e la diffusione non consensuale di *deepfake* che spaziano dalla pornografia da vendetta artificiale (*deepnude*) alla manipolazione informativa finalizzata all'aggiotaggio o alla destabilizzazione politica, il legislatore interno si è trovato dinanzi a un bivio dogmatico: procedere a un'estensione analogica delle fattispecie incriminatrici vigenti o cedere alla tentazione dell'ipertrofia sanzionatoria attraverso l'introduzione di reati-manifesto

[12] MARINUCCI-DOLCINI, *Manuale di diritto penale. Parte generale, Milano,* GIUFFRÈ*, 2023, p. 67 ss.*

di nuovo conio.[13] Il tentativo di ricondurre la manipolazione audiovisiva algoritmica nell'alveo dei tradizionali delitti di falso documentale (artt. 476 ss. c.p.) o della sostituzione di persona (art. 494 c.p.) palesa limiti strutturali insuperabili. Sotto il profilo della falsità in atti, l'immutazione del vero richiede la lesione della genuinità o della veridicità di un supporto che possieda una specifica rilevanza giuridica predeterminata dall'ordinamento; l'immagine o l'audio sintetizzato *ex novo* dall'algoritmo non alterano un "atto" nel senso formalistico inteso dal codice del 1930, bensì creano un simulacro che si colloca al di fuori del perimetro precettivo classico.[14] Del pari, il delitto di cui all'art. 494 c.p. postula l'induzione in errore del terzo circa l'identità fisica del soggetto agente, presupponendo una condotta che mal si concilia con l'astrattezza disincarnata della propagazione virale sui *network* telematici, dove l'autore della condotta manipolativa rimane spesso del tutto anonimo e la lesione si consuma in capo a una platea indistinta di consociati.[15] L'insufficienza degli strumenti classici spiega l'intervento legislativo di urgenza che ha portato all'introduzione di specifiche fattispecie volte a reprimere la diffusione illecita di contenuti generati tramite intelligenza artificiale.[16] Tuttavia, un'analisi rigorosa di queste disposizioni evidenzia un evidente *deficit* di determinatezza strutturale. La descrizione del mezzo tipico della condotta sovente ancorata a formule elastiche quali "sistemi tecnologici idonei a indurre in errore sulla genuinità del contenuto" finisce per delegare al giudice di merito, e inevitabilmente al suo perito d'ufficio, l'esatta determinazione del discrimine tra il penalmente rilevante e il penalmente lecito.[17] Si assiste così a una pericolosa "tecnicizzazione della norma penale", in cui l'elemento oggettivo del reato non è delineato con chiarezza dal legislatore, ma dipende dallo stato dell'arte tecnologico sussistente al momento del fatto, con palese *vulnus* del principio di calcolabilità

[13] CUPELLI, *La legalità penale alla prova della modernità tecnologica,* in *Dir. pen. contemp., 2021, n. 2, p. 45.*

[14] PALAZZO, *I confini della tutela penale tra evoluzione tecnologica e tutele costituzionali,* in *Rivista italiana di diritto e proc.pen, 2019, p. 15 ss.*

[15] CANESTRARI, *I delitti contro la persona nell'era digitale,* in *Diritto Penale e Processo, 2022, fasc. 5, p. 589.*

[16] Il riferimento è all'introduzione di fattispecie incriminatrici di matrice settoriale, quali l'art. 612-*ter* c.p. in materia di diffusione illecita di immagini o video sessualmente espliciti, e ai più recenti progetti di riforma sanzionatoria legati all'impiego illecito di sistemi di intelligenza artificiale.

[17] PADOVANI, *dir. pen., Milano, Giuffrè, 2023, p. 134.*

giuridica e di conoscibilità del precetto.[18] Il deficit di tassatività e determinatezza strutturale che affligge le risposte penalistiche tradizionali dinanzi alla fenomenologia dei *deepfake* emerge con plastica evidenza laddove si sposti l'indagine sul versante della tutela dei beni giuridici della personalità, con specifico riguardo alla dignità umana e alla libertà sessuale. L'archetipo di tale frizione dogmatica è rappresentato dal delitto di diffusione illecita di immagini o video sessualmente espliciti di cui all'art. 612-*ter* c.p.[19], una fattispecie originariamente plasmata dal legislatore sul presupposto della captazione o della sottrazione di materiale "reale", ossia documentante un atto sessuale effettivamente compiuto dal soggetto passivo.[20] L'avvento delle tecnologie di intelligenza artificiale generativa e, nello specifico, delle applicazioni di *deepnude* basate su architetture *GAN* introduce un elemento di radicale discontinuità che cortocircuita l'alveo precettivo della norma. Nel momento in cui il reo realizza e diffonde in rete un filmato sessualmente esplicito sovrapponendo il volto della vittima sul corpo di un'attrice pornografica, ci si trova dinanzi a un simulacro disincarnato, a una realtà simulata in cui il soggetto passivo non ha mai prestato il proprio corpo né ha mai compiuto l'atto rappresentato.[21] Sotto il profilo dogmatico, siffatta condotta svela l'inadeguatezza dell'art. 612-*ter* c.p. rispetto al principio di legalità (art. 25, co. 2, Cost.). Se l'interprete qualifica come "immagini sessualmente esplicite" della vittima un dato digitale interamente sintetizzato *ex novo* dall'algoritmo, rischia di compiere un'operazione di estensione analogica in *malam partem*, sanzionando la condotta sulla base dell'effetto visivo finale anziché sulla materialità del fatto tipico[22]. Non si ignora quella pur lodevole supplenza interpretativa della giurisprudenza di merito che, mossa da comprensibili istanze di tutela della vittima, tende a superare

---

[18] MANTOVANI, *Diritto penale. Parte generale, Padova, CEDAM, 2023, p. 212.*

[19] *I*ntrodotto dall'art. 1, co. 1, L. 19 luglio 2019, n. 69 (c.d. Codice Rosso). Per un primo inquadramento della fattispecie, CADOPPI, *Il Codice Rosso: commento legge per legge, Milano, Giuffrè, 2020, p. 112 ss.*

[20] Sul concetto di «materiale sessualmente esplicito» e sulla necessità che esso ritragga una condotta reale ai fini della tipicità, FERRANDO, *La tutela penale della sfera sessuale nell'era digitale,* in *Rivista Italiana di Dir. proc. pen., 2021, fasc. 3, p. 945.*

[21] MANES, *L'orizzonte artificiale del diritto penale: l'impatto dell'IA sul sistema dei reati,* in *Dir. pen. contemp. Rivista Trimestrale, 2022, n. 2, p. 14 ss.,* il quale evidenzia come la sintesi biometrica totale scardini il concetto classico di corpo come supporto materiale del reato.

[22] Sul divieto di analogia in *malam partem* legato all'evoluzione tecnologica, il classico PALAZZO, *I confini della tutela penale: legalità e interpretazione, saggio in onore di??? Milano, Giuffrè, 2019, p. 78.* In termini analoghi, FLORA, *Il principio di determinatezza e i nuovi media, Padova, CEDAM, 2023, p. 202.*

l'*impasse* ravvisando l'oggetto della protezione oggettiva non già nel compimento materiale dell'atto sessuale, bensì nel grave e identico *vulnus* arrecato alla dignità e alla reputazione sociale del soggetto la cui identità visiva sia stata strumentalizzata.[23] Tuttavia, per quanto l'esigenza di protezione sia sacrosanta, la gravità del danno fenomenico non può legittimare il sacrificio del principio di tassatività e determinatezza (art. 25, co. 2, Cost.) sull'altare dell'emergenza tecnologica. La sanzione penale esige la rigorosa corrispondenza tra fatto tipico e condotta materiale; l'estensione del perimetro applicativo ai simulacri interamente sintetizzati dall'algoritmo spetta unicamente al legislatore, mediante la formulazione di una fattispecie autonoma che non forzi i confini semantici del dettato normativo vigente. Dal punto di vista del bene giuridico protetto, si consuma una vera e propria mutazione dell'oggetto dell'offesa: non viene più lesa la *privacy* o la riservatezza della vita sessuale (giacché quell'atto non appartiene alla biografia della vittima), bensì viene aggredita l'identità personale e l'onorabilità sociale nella sua dimensione più intima.[24] L'insufficienza dei rimedi vigenti non è emendabile attraverso l'introduzione di fattispecie speciali modellate sulla falsariga dei reati-manifesto, le quali nell'ansia di rincorrere l'evoluzione tecnologica ricorrono a clausole elastiche incentrate sull'idoneità dei mezzi ingannevoli, finendo per demandare l'accertamento penale alle fluttuazioni dello stato dell'arte informatico.[25] Si delinea così il paradosso di una norma penale strutturalmente incompleta, la cui applicazione dipende dal verdetto del consulente tecnico, con grave *vulnus* per la calcolabilità giuridica e per il diritto del consociato di conoscere preventivamente i confini del penalmente lecito.[26] Dalla dimensione micro-individuale dei diritti della personalità, la forza destabilizzante del falso algoritmico si riverbera simmetricamente sulle macro-categorie dei beni giuridici collettivi, lambendo la pubblica fede, la trasparenza dei mercati finanziari e la stessa stabilità delle istituzioni

[23] Sul principio di tassatività e sul divieto di analogia in *malam partem* quali corollari del principio di legalità penale, v. Corte cost., n. 115 del 2018 e Corte cost., n. 98 del 2021, che ribadisce l'impossibilità di colmare per via interpretativa lacune rimesse alla discrezionalità del legislatore.

[24] Sulla scomposizione del bene giuridico nei reati informatici e sulla transizione dalla *privacy* all'identità digitale, MUCCIARELLI, *I reati contro la persona telematici,* in *Trattato di Dir. pen.,* diretto da CINGARI, *Torino,* GIAPPICHELLI, *2024, p. 310 ss.*

[25] PULITANÒ, *Politica criminale e reati-manifesto,* in *Criminalia, 2020, p. 45.* L'Autore stigmatizza la tendenza del legislatore contemporaneo a creare fattispecie «simboliche» prive di reale determinatezza precettiva.

[26] Sulla «tecnicizzazione» della norma penale e la conseguente delega al sapere peritale, DONINI, *Il diritto penale giurisprudenziale e i saperi scientifici,* in *Dir. pen. e Processo, 2021, p. 1120.*

democratiche. Il potenziale offensivo risiede, ancora una volta, nella mutazione ontologica della prova operata dalle reti neurali: il *deepfake* non si limita a veicolare una notizia falsa, ma fabbrica un fatto storico visivamente inconfutabile, capace di generare un impatto cognitivo immediato e irreversibile sulla collettività. Si consideri lo scenario in cui venga diffuso capillarmente sui circuiti telematici un video sintetico del Governatore della Banca centrale che annunci un imminente e catastrofico *default* del sistema bancario, ovvero del Presidente del Consiglio che dichiari lo stato di mobilitazione militare. Dinanzi a una siffatta propagazione virale, l'apparato precettivo classico rivela limiti strutturali insuperabili. La fattispecie di aggiotaggio (art. 2637 c.c. e art. 501 c.p.), incentrata sulla diffusione di "notizie false" o sull'impiego di "altri artifici", fatica a inglobare la complessità di un fenomeno in cui l'artificio non è una condotta dissimulatoria che accompagna il dato, ma è il dato stesso, immanente alla struttura biometrica del *file*. Inoltre, il delitto di procurato allarme presso l'Autorità (art. 658 c.p.) si rivela del tutto inidoneo a sanzionare l'offensività globale dell'evento, venendo declassato a mera contravvenzione a fronte di condotte in grado di paralizzare l'ordine pubblico o di bruciare miliardi di capitali sui mercati regolamentati in pochi secondi. La scelta legislativa di anticipare la soglia di punibilità alla tutela del pericolo astratto, lungi dal costituire un efficace argine sanzionatorio, si scontra con il rischio latente di un'ipertrofia sanzionatoria liberticida. Formule penali onnicomprensive, volte a reprimere genericamente l'inedita fenomenologia della disinformazione algoritmica, rischiano infatti di flettere la norma penale verso indebite derive di controllo e censura, capaci di comprimere tanto le dinamiche del dissenso politico quanto le libere manifestazioni dell'espressione artistica, in aperto contrasto con le guarentigie costituzionali scolpite dall'art. 21 della Carta costituzionale.[27] Tuttavia, l'offensività del falso algoritmico non esaurisce i suoi riverberi sul piano del diritto sostanziale; al contrario, il fulcro dell'impatto destabilizzante delle tecnologie di intelligenza artificiale generativa si colloca nel cuore pulsante della dialettica processuale, ridefinendo i parametri di operatività del principio del «libero convincimento del giudice» (ex art. 192, co. 1, c.p.p.) e la struttura logica della regola di giudizio codificata all'art. 533, co. 1,

[27] Sul punto, l'ampia disamina di MANES, *L'oracolo algoritmico e la giustizia penale*, Bologna, Il Mulino, 2025, p. 98.

c.p.p.[28] Nella tassonomia codicistica tradizionale, il documento fonografico o cinematografico ha sempre beneficiato di una presunzione implicita di "immediatezza rappresentativa", tale da relegare l'attività di verifica del giudicante a un mero vaglio di pertinenza e di non macroscopica alterazione materiale.[29] Come già anticipato, la diffusione dei *deepfake* incrina il tradizionale rapporto di immediata corrispondenza tra documento audiovisivo e fatto rappresentato. Ne consegue che l'evidenza digitale può richiedere essa stessa un autonomo accertamento circa la propria provenienza e genuinità. La ricaduta più insidiosa investe l'onere probatorio, determinando il rischio di una sua strisciante e surrettizia inversione a danno della difesa.[30] Di fronte alla contestazione difensiva che eccepisca la natura parzialmente o totalmente manipolata di un supporto audiovisivo o informatico telematico, il sistema processuale rischia l'impasse. Se la giurisprudenza di legittimità tende a liquidare le allegazioni difensive prive di specifico supporto tecnico come "mere congetture" inidonee a incrinare il compendio probatorio, la natura intrinsecamente impercettibile delle manipolazioni operate da reti neurali generative evolute rende impossibile per l'imputato fornire la prova positiva della falsificazione.[31] Si profila così una complessa linea di faglia nell'architettura dell'art. 533 c.p.p. Affinché un dubbio possa qualificarsi come "ragionevole", esso deve scaturire da elementi razionali, obiettivi e non meramente ipotetici. Tuttavia, nel momento in cui la scienza forense attesta l'assoluta e indistinguibile fungibilità tra realtà materiale e simulazione algoritmica, l'ipotesi della manipolazione cessa di essere un'astratta elucubrazione per tramutarsi in un'alternativa dotata di intrinseca plausibilità scientifica.[32] Per disinnescare lo stallo epistemologico e l'effetto del *liar's dividend*, si rinvia al modello dell'onere qualificato di allegazione a carico della difesa già ampiamente delineato nel primo paragrafo. Fuori da questa direttrice, il giudice si trova privato della sua tradizionale po-

[28] TONINI-CARLIZZI, *Manuale di procedura penale, Milano, Giuffrè, 2023, p. 245 ss.;* nonché RAFARACI, *La prova scientifica nel processo penale tra canoni di valutazione e garanzie difensive,* in *Rivista italiana di diritto e procedura penale, 2019, fasc. 2, p. 545 ss.*

[29] ILLUMINATI, *La disciplina delle prove,* in CONSO-GREVI-BARGIS, *Compendio di procedura penale,* Padova, CEDAM, 2023, p. 412.

[30] FERRUA, *La prova nel processo penale, Torino, Giappichelli, 2019, p. 118.*

[31] CONTI, *Scienza e processo penale: nuove sfide e antiche garanzie,* in *Cassazione Penale, 2022, p. 1450 ss.*

[32] DANIELE, *Regole di giudizio e prove scientifiche: il giudice di fronte all'invisibile,* in *Rivista di diritto processuale, 2024, p. 889.*

stura di *peritus peritorum*: egli non è più chiamato a valutare l'attendibilità di una testimonianza o la logicità di un'inferenza indiziale, ma diviene ostaggio di un conflitto tra consulenze tecniche, i cui protocolli di analisi sono connotati da margini di errore statistico ineliminabili e da una congenita opacità metodologica. Delineata la faglia profonda che incrina la tenuta dei canoni probatori classici, emerge la necessità di abbandonare l'approccio diagnostico "a posteriori" storicamente affidato alla perizia tecnologica *ex* art. 220 c.p.p. per inaugurare una strategia di intervento normativo incentrata su filtri di ammissibilità strutturali e "a monte". La soluzione ricostruttiva non può tuttavia cedere a utopie tecnocratiche: l'introduzione di un obbligo generalizzato di marcatura crittografica all'origine rischierebbe di mutilare il compendio probatorio, espellendo dal processo la quasi totalità delle informazioni genuine prodotte spontaneamente dai privati.[33] Al fine di superare tale aporia strutturale, si prospetta *de jure condendo* l'introduzione di un modello normativo fondato su un "doppio binario" di ammissibilità della prova digitale, da codificarsi mediante una riforma coordinata e sinergica degli artt. 234 e 234-*bis* c.p.p. Il primo binario, deputato alla disciplina della c.d. evidenza digitale nativa-investigativa, individua quale oggetto dati conoscitivi generati direttamente dagli organi inquirenti o dalla polizia giudiziaria nell'esercizio delle funzioni tipiche, quali le captazioni ambientali, i tracciamenti satellitari o le videoriprese. Per siffatta categoria di prove, l'adozione di rigorosi protocolli di autenticazione crittografica mediante tecnologie di registri distribuiti, firme digitali e funzioni di *hashing* contestuali all'atto della captazione del dato deve essere prescritta a pena di inutilizzabilità patologica ai sensi dell'art. 191 c.p.p. Lo Stato, in quanto detentore del monopolio e del controllo esclusivo della sorgente tecnologica, non può infatti giovarsi di risultanze probatorie di cui non sia istituzionalmente in grado di garantire l'assoluta e oggettiva incorruttibilità originaria.[34]

---

[33] LUPARIÀ, *L'evidenza digitale nel processo penale italiano, Milano, Giuffrè, 2021, p. 174.*

[34] In ordine alle potenzialità applicative della tecnologia blockchain e dei protocolli crittografici alla garanzia di integrità dell'evidenza informatica, si veda SIGNORATO, *Le indagini digitali. I mezzi di ricerca della prova informatica nel processo penale, Torino,* GIAPPICHELLI, *2020, p. 112 ss.;* in chiave pionieristica, *altresì* ZICCARDI, *L'investigazione digitale informatica tra acquisizione della prova e rispetto dei diritti fondamentali,* in *Rivista italiana di diritto e proc. pen., 2017, fasc. 3, p. 945 ss. S*ulla necessità di standardizzare la catena di custodia nello spazio giuridico europeo, prescindendo da asettici schemi formalistici, CAIANIELLO, *Premesse per una teoria della prova nel processo penale europeo,* in *Rivista italiana di diritto e proc. pen., 2022, fasc. 3, p. 911.*

Sotto il profilo operativo e sanzionatorio, d'altronde, l'evocata inutilizzabilità patologica ex art. 191 c.p.p. deve essere calibrata con realismo strutturale, onde evitare la paralisi delle funzioni inquirenti dinanzi alle iniziali asimmetrie tecnologiche territoriali della polizia giudiziaria. In una prospettiva di transizione graduale, l'assenza di una marcatura crittografica nativa da parte dell'apparato statale non deve comportare l'inutilizzabilità *tout court* del dato, bensì escludere l'immediata operatività di qualsiasi presunzione di conformità o genuinità dell'atto, imponendo una verifica in concreto della traccia. Ne deriva l'integrale e non surrogabile traslazione dell'onere probatorio in capo al Pubblico ministero, il quale sarà gravato dall'obbligo di ricostruire e dimostrare la purezza della catena di custodia con standard di eccezionale *rigor*, senza poter beneficiare di alcuna semplificazione istruttoria. Il secondo binario, di contro, perimetra lo statuto della prova digitale nativa-privata, concernente i documenti informatici formati al di fuori del contesto investigativo ed entrati nel procedimento *aliunde*, come i supporti video registrati da dispositivi mobili o i flussi di comunicazioni telematiche tra consociati. In tale ipotesi, stante l'impossibilità oggettiva di pretendere una marcatura crittografica alla sorgente, qualora la difesa formuli una contestazione specifica e dotata di intrinseca plausibilità scientifica, la prova audiovisiva perde la propria ordinaria valenza di prova rappresentativa autosufficiente. Essa subisce una vera e propria *capitis deminutio* sul piano epistemologico. Muovendo, d'altronde, dall'orizzonte del diritto vigente, la traccia priva di elementi di autenticazione crittografica sconta una strisciante retrocessione dal rango di prova documentale diretta a quello di mero elemento indiziario, soggetto come tale al canone valutativo del rigido concorso di elementi precisi e concordanti ai sensi dell'art. 192, comma 2, c.p.p. Anche nell'attuale quadro dogmatico, il dato rappresentato il dato rappresentato sul supporto telematico non potrà fondare, in via esclusiva, il giudizio di colpevolezza, gravando sul giudicante l'obbligo di riscontrarlo rigorosamente attraverso un compendio probatorio integrativo di natura analogica. È doveroso precisare, in chiave di rigoroso rispetto del principio di tassatività e per scongiurare derive di prova legale strisciante, che tale retrocessione epistemica della prova digitale non crittografata al rango di mero indizio non si realizza, nel sistema vigente, in via automatica o aprioristica. Una simile rigidità contrasterebbe con il principio cardine del libero convincimento del giudice e con il rifiuto storico di schemi probatori precostituiti dall'ordinamento. Pertanto, la *capitis deminutio* dell'evidenza audiovisiva privata deve

configurarsi esclusivamente quale *extrema ratio*, destinata a scattare laddove i protocolli forensi di *detection standard* restituiscano esiti probabilistici insanabilmente incerti, o algoritmi di segno opposto lascino permanere un dubbio scientifico oggettivo e non altrimenti superabile circa la genuinità della traccia. Tale compendio dovrà basarsi sull'esame testimoniale in contraddittorio ai sensi dell'art. 111 Cost., integrato da prove materiali oggettive. Il giudice sarà così chiamato a una valutazione complessiva della coerenza logico-comportamentale delle parti, verificando che la ricostruzione dei fatti trovi riscontro nella reale dinamica del movente.[35] Un siffatto modello trova un solido addentellato sistematico nell'evoluzione del quadro regolatorio sovranazionale, con particolare riferimento alla disciplina degli ordini europei di produzione e conservazione delle prove elettroniche, nonché ai requisiti di conformità imposti dal Regolamento UE sull'Intelligenza Artificiale ai sistemi ad alto rischio impiegati nel settore del *law enforcement*.[36] L'estensione di tali *standard* al rito interno consentirebbe di disinnescare i profili di criticità sottesi al "ragionevole dubbio algoritmico": l'esibizione di una catena di custodia digitale certificata sin dall'origine esclude in radice l'ipotesi di una manipolazione successiva mediante reti neurali generative, restituendo alla dialettica processuale un perimetro di certezza oggettiva su cui innestare il libero convincimento del giudice.[37] Non si tratta di reintrodurre surrettiziamente un anacronistico regime di prova legale, bensì di perimetrare razionalmente l'oggetto del giudizio, impedendo che l'accertamento penale si converta in una simulazione virtuale in cui la verità materiale soccombe dinanzi all'estetica del verosimile. In conclusione, la risposta dogmatica all'insidioso dilemma dei *deepfake* non può esaurirsi nel mero inseguimento tecnologico di sofisticati moduli di *deepfake detection*, strutturalmente destinati a essere superati dalle

---

[35] Sulla parabola evolutiva della prova documentale e sui rischi di una acritica ricezione del dato informatico nel processo penale, si veda l'ampia e fondamentale disamina di CAPRIOLI, *La prova scientifica,* in UBERTIS-CAPRIOLI, *Dir. proc. pen. Vol. II: Atti, prove, provvedimenti, procedimenti, Torino,* GIAPPICHELLI, *2023, p. 115 ss.*

[36] *Regolamento (UE) 2023/1543 del Parlamento europeo e del Consiglio, del 12 luglio 2023,* relativo agli ordini europei di produzione e di conservazione di prove elettroniche nei procedimenti penali. Sul punto, per un quadro d'insieme sulle nuove linee evolutive della cooperazione giudiziaria sovranazionale, KOSTORIS, *Il processo penale nello spazio giuridico europeo, 4ª ed., Torino,* GIAPPICHELLI, *2023, p. 185 ss.;* specificamente sul Regolamento (UE) 2023/1543, si veda CAMALDO, *I nuovi strumenti di cooperazione per l'acquisizione della e-evidence nello spazio giuridico europeo,* in *Dir. pen. e Processo, 2024, fasc. 2, p. 215 ss.*

[37] FERRUA, *Il processo penale: profilo teorico e sistematico, Torino,* GIAPPICHELLI, *2022, p. 302.*

successive generazioni di reti neurali. Al contrario, la soluzione impone di riaffermare l'antropocentrismo del processo penale. Il canone dell'«oltre ogni ragionevole dubbio» conserva la propria tenuta epistemica dinanzi alle derive della post-verità solo se il giudicante ricolloca il dato digitale all'interno di una dimensione olistica e storicizzata del giudizio, nella consapevolezza che la sanzione penale incide sulla libertà inviolabile dell'uomo (*ex* art. 13 Cost.) e che nessuna proiezione algoritmica, per quanto biologicamente perfetta, potrà mai abdicare alla complessa, faticosa e squisitamente umana ricerca della verità materiale.[38]

3. La dimensione sovranazionale e il governo tecnologico del processo. Il modello interpretativo del "doppio binario", volto a preservare la tenuta del canone dell'oltre ogni ragionevole dubbio nell'orizzonte della *infocalypse* audiovisiva, trova un decisivo punto di emersione e di sponda nel quadro regolatorio europeo di recente introduzione. Il punto di riferimento obbligato è costituito dal Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio, del 13 giugno 2024, stabilente regole armonizzate sull'intelligenza artificiale (noto come *Artificial Intelligence Act* o *AI Act*).[39] L'impianto normativo eurounitario, strutturato attorno a un approccio basato sul rischio (*risk-based approach*), inserisce espressamente i sistemi di *AI* destinati ad essere utilizzati dalle autorità di contrasto (*law enforcement*) nell'allegato III, qualificandoli come sistemi ad "alto rischio" (*high-risk AI systems*).[40] La qualificazione in termini di alto rischio comporta l'insorgenza di stringenti obblighi di conformità in capo ai fornitori e ai soggetti utilizzatori (*deployers*), tra i quali spiccano il controllo umano, la robustezza tecnica, la sicurezza informatica e, per quanto qui maggiormente rileva, i doveri di trasparenza e tracciabilità del-

[38] MANES, *l'oracolo algoritmico e la giustizia penale, Bologna, Il Mulino, 2025, p. 185*

[39] *Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio, del 13 giugno 2024,* stabilente regole armonizzate sull'intelligenza artificiale (regolamento sull'intelligenza artificiale). *Per un primo inquadramento penalistico delle nuove regole europee, si veda AMATI, I confini penali dell'Intelligenza Artificiale nell'orizzonte europeo, Milano,* GIUFFRÈ, 2024, p. 45 ss.

[40] V., in particolare, il punto 6 dell'Allegato III del Regolamento, relativo ai sistemi di *AI* destinati ad essere utilizzati dalle autorità di contrasto per effettuare valutazioni dei rischi, profilazione o per individuare "falsi documentali" *v.* CAIANIELLO, *Algoritmizzazione della giustizia penale e tutele sovranazionali: l'impatto dell'AI Act,* in *Diritto Penale e Processo, 2024, n. 9, p. 1120 ss.*

le operazioni (artt. 12 e 13 Reg. UE 2024/1689. Sotto il profilo strettamente processuale, siffatti requisiti di fonte sovranazionale si scontrano frontalmente con l'attuale prassi forense della *deepfake detection*. Gli algoritmi deputati all'accertamento diagnostico della natura sintetica o manipolata di un *file* (siano essi basati sull'analisi delle asincronie biometriche o sull'individuazione di artefatti microscopici generati dalle reti antagoniste generative GAN) sono, nella quasi totalità dei casi, *software* commerciali "proprietari", protetti da privativa industriale o da segreto commerciale (*trade secret*).[41] Si profila in tal modo un inedito e profondo punto di frizione con le garanzie fondamentali scolpite nell'art. 111 Cost. e nell'art. 6 C.E.D.U. Sul piano dell'epistemologia giudiziaria, l'avvento dell'*AI* generativa segna il definitivo crollo dei paradigmi consolidati della *digital forensics* classica. Se l'informatica forense tradizionale fondava l'autenticità del dato sul rigido rispetto della catena di custodia e sulla verifica dell'integrità del supporto tramite stringhe di *hash*, la fenomenologia dei *deepfake* azzera siffatte certezze: un documento cross-mediale totalmente artificiale può presentare un'impronta informatica perfettamente integra e una tracciabilità di sequestro impeccabile, pur essendo ontologicamente falso. La computazione algoritmica di *detection*, pertanto, non è più chiamata a verificare l'alterazione storica del reperto, bensì l'autenticità genetica della sua stessa manifestazione materiale.[42] Nel momento in cui l'accertamento di una prova informatica decisiva ai fini della responsabilità penale poggia sui risultati forniti da un *software* di *detection*, il diritto di difesa (art. 24 Cost.) e il principio del contraddittorio nella formazione della prova subiscono una drastica contrazione qualora il codice sorgente (*source code*) e i dati di addestramento (*training data*) dell'algoritmo rimangano preclusi allo scrutinio delle parti. La tutela del segreto industriale delle *software house* non può tradursi nell'ingresso all'interno del circuito processuale di una scatola nera (*black box*), il cui *output* sia sottratto al controesame metodologico della difesa. Un simile sbarramento informativo determina un *vulnus* intollerabile alla struttura dell'art.

[41] Sul delicato bilanciamento tra proprietà intellettuale e diritto alla prova nell'era digitale, v. QUATTROCCOLO, Artificial *Intelligence, Computational Modelling and Criminal Proceedings, Springer, 2020, p. 73 ss.;* v. altresì BONINI, *La prova scientifica digitale e il segreto commerciale: nuove barriere per il contraddittorio,* in *Rivista Italiana di Diritto e Procedura Penale, 2023, fasc. 2, p. 580 ss.*

[42] *S*ul superamento metodologico della digital forensics classica al cospetto delle manipolazioni da intelligenza artificiale generativa, si veda l'approfondita analisi di COSTABILE, *Metodologie di accertamento della prova informatica: dalle stringhe di hash alla verifica del falso algoritmico,* in *Giurisprudenza Penale, 2024, n. 4, p. 310 ss.*

111, co. 5, Cost., risolvendosi nell'introduzione surrettizia di una sorta di "testimonianza tecnologica anonima". Se alla difesa è inibito l'accesso ai criteri che governano la decodificazione del dato, il diritto di "esaminare o fare esaminare" i supporti d'accusa si converte in un simulacro, poiché l'elaborato della macchina sostituisce la deposizione del perito, sottraendosi alla dialettica del controesame metodologico.[43] È proprio in questo crinale che si radica il rischio di una "devoluzione tecnologica" della funzione giudicante, amplificato dal fenomeno dell'*automation bias*: l'inclinazione psicologica del giudice a fidarsi acriticamente del responso computazionale, percepito come oggettivo e privo di fallibilità. La giurisprudenza di legittimità, pur essendosi pronunciata in tema di captatore informatico (*trojan horse*) salvaguardando parzialmente la riservatezza delle metodologie operative della polizia giudiziaria, non ha ancora affrontato *ex professo* il tema dell'opacità degli algoritmi valutativi applicati alla prova documentale.[44] Occorre d'altronde evidenziare che i sistemi di *detection* non restituiscono certezze assolute, bensì meri calcoli probabilistici condizionati da tassi intrinseci di errore statistico (falsi positivi e falsi negativi). L'inesprimibilità del dato in termini di certezza assoluta implica che il responso algoritmico rechi in sé, strutturalmente, la cifra del "ragionevole dubbio" (*ex* art. 533, co. 1 c.p.p.), imponendo al decidente di rifiutare qualsiasi equiparazione tra verdetto tecnologico e pieno accertamento del fatto. Se l'algoritmo opera come una *black box* inaccessibile, il giudizio di colpevolezza si converte in una acritica ratifica di un responso *machine-to-machine*, svuotando di significato la logica del libero convincimento (*ex* art. 192, co. 1 c.p.p.).[45] Nel tentativo di delineare un quadro sistematico, occorre precisare che *l'AI Act* possiede una natura eminentemente amministrativo-regolatoria: esso disciplina la conformità del prodotto sul mercato, ma non norma gli effetti processuali derivanti dall'impiego di uno strumento opaco. La tesi che

[43] RUGGERI, *Giusto processo e intelligenza artificiale: la tutela dei diritti fondamentali tra algoritmi predittivi e scatole nere,* in *Dir. pen. contemp., 2023, n. 1, p. 45...* Sul concetto di opacità algoritmica applicata al rito penale, v. l'ampia e fondamentale ricostruzione di PAONESSA, *Il processo penale nell'era degli algoritmi. Conoscenza e garanzie tra epistemologia e diritto, Torino,* GIAPPICHELLI, *2022.*

[44] Sull'evoluzione giurisprudenziale in materia di prove tecnologiche e limiti del segreto, cfr. *Cass. pen., Sez. un., 1° luglio 2016, n. 26889,* in *Cassazione Penale, 2016, p. 3850* GALIZIA, *La prova scientifica e l'algoritmo: l'orizzonte mobile del contraddittorio,* in *Rivista Italiana di Diritto e Proc.pen, 2024, fasc. 1, p. 112 ss.*

[45] Sulla valenza esclusivamente probabilistica degli output algoritmici forensi e sul dovere del giudice di sottoporre il dato a riscontri estrinseci NEGRI, *La logica della prova scientifica probabilistica nel prisma dell'art. 533 c.p.p.,* in *Diritto Proc. pen., 2024, n. 2, p. 185 ss.*

qui si sostiene è che le previsioni del Regolamento europeo debbano assurgere a "parametro interposto" di attendibilità scientifica per il giudice penale. Un *software* di *detection* privo della documentazione tecnica richiesta dall'art. 13 dell'*AI Act* deve essere ritenuto *ipso facto* privo dei requisiti minimi di affidabilità, con conseguente inammissibilità della relativa consulenza tecnica per manifesta inidoneità metodologica. Al fine di contemperare la tutela della proprietà intellettuale delle *software house* con il diritto al contraddittorio, la soluzione interpretativa non può risolversi nel rigetto aprioristico dei sistemi commerciali, né in una cieca capitolazione di fronte al segreto industriale. Piuttosto, occorre mutuare modelli di accesso controllato, sul paradigma degli accertamenti tecnici guidati da stringenti protocolli di *disclosure* parziale. Se il cuore del brevetto (il codice sorgente nativo) può godere di forme di protezione riservata, deve ritenersi precondizione inderogabile di ammissibilità della prova l'ostensibilità totale dei dati di addestramento (*training data*) e delle metriche di errore statistico. Solo consentendo alla difesa di effettuare una *cross-validation* indipendente dell'algoritmo sottoponendolo a *dataset* di controllo speculari è possibile restituire all'organo giudicante l'effettivo governo del sapere scientifico, emancipandolo da una supina ricezione del responso tecnologico e salvaguardando lo statuto epistemologico del "giusto processo.[46] Il superamento dell'opacità tecnologica della *black box*, mediante protocolli di *disclosure* controllata e *cross-validation* dei dati di addestramento, si scontra inevitabilmente con la dimensione intrinsecamente transnazionale della prova digitale forense. Nel prisma dell'*infocalypse* audiovisiva, l'accertamento del fatto reato non è quasi mai un'operazione circoscritta entro i confini dello Stato sovrano: i reperti multimediali artificiali (video *deepfake*, audio clonati, flussi di disinformazione algoritmica) vengono generati tramite infrastrutture *cloud* distribuite e veicolati attraverso piattaforme di comunicazione gestite da *Service Providers* situati in giurisdizioni estere, prevalentemente *extra*-UE.[47] In questo scenario, lo strumento normativo destinato a ridefinire le coordinate dell'acquisizione probatoria è il c.d. Pacchetto *e-Evidence*, costituito dal Rego-

[46] Sulla necessità di elaborare nuovi protocolli di disclosure e di validazione scientifica della prova algoritmica forense, ORLANDI, *Epistemologia giudiziaria e intelligenza artificiale. Il controllo di scientificità della prova cibernetica,* in *Processo penale e Giustizia, 2024, n. 3, p. 495 ss.*

[47] *Regolamento (UE) 2023/1543 del Parlamento europeo e del Consiglio, del 12 luglio 2023,* relativo agli ordini europei di produzione e di conservazione di prove elettroniche nei procedimenti penali, in *GUUE, L. 191, 28 luglio 2023.*

lamento (UE) 2023/1543 e dalla Direttiva (UE) 2023/1544, volti ad istituire gli ordini europei di produzione e di conservazione delle prove elettroniche nelle materie penali.[48] L'innovazione macro-sistemica introdotta dal legislatore eurounitario risiede nel definitivo superamento del tradizionale schema intergovernativo della rogatoria internazionale e dell'Ordine europeo di indagine (OEI), sostituiti da un meccanismo di interlocuzione diretta *authority-to-provider*. L'autorità giudiziaria dello Stato di emissione (nel sistema italiano, il Pubblico ministero o il giudice) può ingiungere direttamente al rappresentante legale del prestatore di servizi stabilito nell'Unione di produrre i dati informatici protetti, prescindendo dalla mediazione dello Stato di esecuzione.[49] Sotto il profilo del rigore metodologico e dell'analisi dei rischi richiesti da una rigorosa disamina scientifica, siffatta accelerazione procedurale introduce tensioni latenti di estrema gravità rispetto al principio di uguaglianza delle armi (*ex* art. 6 C.E.D.U.) e al diritto di difesa (*ex* art. 24 Cost.). Se l'organo dell'accusa acquisisce direttamente dal colosso tecnologico d'oltreoceano il file manipolato o i metadati di tracciamento dell'attacco algoritmico, la difesa si trova confinata in una posizione di radicale asimmetria informativa. Il Regolamento (UE) 2023/1543 attribuisce i rimedi giurisprudenziali avverso l'ordine di produzione principalmente alla persona i cui dati sono richiesti e al *provider* stesso (art. 17), ma omette di disciplinare uno statuto difensivo *endo*-processuale che consenta all'imputato di interloquire *ex ante* sulle modalità di estrazione e di preservazione dell'integrità del dato nativo.[50] La soluzione interpretativa che in questa sede si propone, al fine di offrire un contributo ricostruttivo originale e sistematico, consiste nel configurare un filtro di ammissibilità sbarrante fondato sul combinato disposto tra l'art. 191 c.p.p. e le garanzie dell'art. 111 Cost. Qualora l'acquisizione transnazionale del documento informatico tramite ordine di produzione europeo avvenga senza la contestuale acquisizione della documentazione tecnica relativa ai protocolli di estrazio-

[48] Sul punto, per un quadro d'insieme sulle nuove linee evolutive della cooperazione giudiziaria sovranazionale, KOSTORIS, *Il processo penale nello spazio giuridico europeo*, 4ª ed., Torino, GIAPPICHELLI, 2023, p. 185 ss.

[49] Sullo statuto epistemologico della prova e sulla scomposizione logica del giudizio di colpevolezza nell'orizzonte tecnologico, si veda l'opera fondamentale di UBERTIS, *Elementi di epistemologia giudiziaria, Milano,* Giuffrè*, 1994, p. 55 ss.*

[50] Sul concetto di copia-specchio e sulla ripetibilità teorica dell'estrazione forense tradizionale, v. la celebre ricostruzione di TARUFFO, *La semplice verità. Il giudice e la costruzione dei fatti, Bari-Roma, Laterza, 2009.*

ne applicati dal *provider* estraneo, la prova dovrà ritenersi viziata da inutilizzabilità derivata per violazione del diritto al giusto processo. Non è tollerabile l'ingresso nel compendio probatorio di un reperto digitale la cui genuinità genetica non sia verificabile a causa dell'*opacità della catena di custodia transnazionale*.[51] La cooperazione sovranazionale non può tradursi in una franchigia per l'accusa. Affinché la prova algoritmica transnazionale possa legittimamente superare lo sbarramento del giudizio di ammissibilità, l'organo inquirente dovrà attivare i canali dell'*e-Evidence* non solo per l'acquisizione della prova (l'*output* multimediale), ma anche per l'ostensione dei parametri di errore e dei criteri di validazione del *software* forense eventualmente impiegato all'estero per il tracciamento delle reti *GAN* (*Generative Adversarial Networks*). Solo attraverso questo rigoroso allineamento tra strumenti di cooperazione ed epistemologia giudiziaria è possibile scongiurare il rischio che la fluidità del digitale converta il processo penale in un rito inquisitorio transnazionale, privo di effettivo controllo giurisdizionale. L'analisi sin qui condotta impone di misurarsi con l'interrogativo finale, vero e proprio fulcro del presente contributo: in che modo l'avvento dei *deepfake* e l'opacità dei *software* di *detection* rimodellano la struttura logica della regola di giudizio codificata all'art. 533, co. 1 c.p.p.? La risposta non può esaurirsi in una rassegna delle patologie tecnologiche, ma esige una ricostruzione sistematica dello statuto epistemologico della prova penale nell'era dell'*infocalypse* audiovisiva.[52] La tesi che in questa sede si prospetta muove dalla ridefinizione del concetto di "ragionevole dubbio" al cospetto di prove multimediali potenzialmente artificiali. Nella dinamica processuale classica, il documento fonovideo godeva di una presunzione implicita di corrispondenza al *fait historique*, cedendo solo di fronte a macroscopici indizi di contraffazione. Oggi, la generalizzata disponibilità di architetture generative di livello commerciale inverte radicalmente questo paradigma: la possibilità che un supporto digitale sia un artefatto algoritmico non costituisce più un'ipotesi astratta o iperbolica, bensì un rischio concreto, latente e

[51] Questa impostazione estende e adatta i principi espressi dalle Sezioni Unite in materia di accertamenti irripetibili sui supporti informatici: *Cass. pen., Sez. un., 2 dicembre 2021 (dep. 21 dicembre 2021), n. 44320.*

[52] *Regolamento (UE) 2023/1543 del Parlamento europeo e del Consiglio, del 12 luglio 2023,* relativo agli ordini europei di produzione e di conservazione di prove elettroniche nei procedimenti penali, in *GUUE, L 191, 28 luglio 2023.*

immanente a qualsiasi riscontro dematerializzato.[53] Ne discende che la mera allegazione, ad opera della difesa, del dubbio sull'autenticità del file non può essere liquidata come una strategia dilatoria. Al contrario, ogniqualvolta la difesa introduca un "dubbio di genuinità" suffragato da elementi minimi di verosimiglianza quali l'assenza di una catena di custodia nativa o l'inaccessibilità dei metadati del *file* sorgente, l'onere della prova in capo al Pubblico ministero subisce una profonda mutazione. Non basterà più all'accusa dimostrare che il video "rappresenta" l'imputato nell'atto di commettere il reato; l'accusa avrà l'onere stringente di provare l'autenticità ontologica della fonte generativa, dimostrando l'assenza di manipolazioni algoritmiche mediante metodologie di *detection* che superino il vaglio di trasparenza sovranazionale imposto dall'*AI Act*.[54] È in questo esatto snodo che trova piena operatività il modello del "doppio binario" interpretativo qui propugnato. Il primo binario attiene al giudizio di ammissibilità (lo sbarramento formale): il giudice deve escludere *ab initio* dal compendio probatorio qualsiasi reperto multimediale o consulenza tecnica informatica che si fondi su algoritmi coperti da segreto industriale invalicabile o privi di documentazione sui *training data*. Il secondo binario governa invece il giudizio di valutazione della prova legittimamente ammessa: là dove il *software* forense esprima un responso probabilistico di autenticità (es. probabilità del 95%), quel residuo margine d'errore statistico del 5% (falso positivo) non può essere colmato dal prudente apprezzamento del chiarire. Esso incarna, in forma matematica e oggettiva, la cifra del ragionevole dubbio, precludendo l'utilizzazione del solo dato tecnologico ai fini della pronuncia di condanna.[55] In conclusione, l'emancipazione del giudice dall'*automation bias* e dall'illusione del vero algoritmico si realizza non già nel rifiuto della tecnologia, ma nella sua rigorosa sottomissione alle regole del diritto processuale penale. Solo configurando l'ostensibilità dell'algoritmo come precondizione di

[53] Sul punto, per un quadro d'insieme sulle nuove linee evolutive della cooperazione giudiziaria sovranazionale, v. KOSTORIS, *Il processo penale nello spazio giuridico europeo, 4ª ed., Torino,* GIAPPICHELLI, *2023, p. 185 ss.*

[54] Sulla scomposizione logica del giudizio di colpevolezza e sullo statuto epistemologico della prova i cui moduli concettuali restano fondamentali anche per vagliare l'attendibilità del dato tecnologico e la tenuta del ragionevole dubbio si veda l'opera fondamentale di UBERTIS, *Elementi di epistemologia giudiziaria, Milano,* GIUFFRÈ, *1994, p. 55 ss.*

[55] Questa impostazione estende e adatta i principi espressi dalle Sezioni unite in materia di accertamenti irripetibili sui supporti informatici*: Cass. pen., Sez. un., 2 dicembre 2021 (dep. 21 dicembre 2021), n. 44320.*

ammissibilità della prova e interpretando il margine di errore statistico come sbarramento logico alla condanna, l'ordinamento può conservare la tenuta del canone del giusto processo. In caso contrario, la giurisdizione abdicherebbe al proprio mandato costituzionale, convertendo la verità processuale in una acritica ratifica di un verdetto computazionale privato, con irreparabile *vulnus* per le guarentigie della libertà personale.[56] L'operatività del canone dell'oltre ogni ragionevole dubbio, unitamente alle riflessioni sul modello del "doppio binario", non può esaurire la propria portata precettiva nella sola fase del dibattimento. Ai fini di una compiuta analisi sistematica, occorre mappare i rischi di alterazione e le relative tutele nel segmento cronologico in cui la prova digitale viene ad esistenza e si stabilizza: la fase delle indagini preliminari.[57] Al cospetto della fenomenologia dei *deepfake*, l'attività di repertamento e di prima analisi tecnica condotta dall'organo inquirente assume una valenza intrinsecamente anticipatoria degli esiti del giudizio. La fluidità e la deperibilità del dato cross-mediale generato da intelligenza artificiale impongono un'immediata verifica diagnostica sul supporto sequestrato, al fine di scongiurare la progressiva degradazione dei metadati o la sovrascrittura delle tracce microscopiche di computazione algoritmica latente.[58] In questo scenario, assume una centralità strategica il confine normativo tra l'accertamento tecnico ripetibile (*ex* art. 359 c.p.p.) e l'accertamento tecnico non ripetibile (*ex* art. 360 c.p.p.). La prassi investigativa tende sovente a qualificare la *deepfake detection* alla stregua di una mera operazione tecnica ripetibile, sul presupposto che il *file* digitale sia per sua natura riproducibile in copie identiche.[59] La tesi

---

[56] Per riflessioni speculari sulla tutela dei diritti fondamentali dinanzi alle derive dell'automatizzazione decisionale, si rimanda a *AA.VV., Diritto penale e intelligenza artificiale,* a cura di BRUNELLI*, Milano,* GIUFFRÈ*, 2022.*

[57] Sulla centralità delle indagini preliminari nella formazione della prova scientifica digitale e sulla necessità di un innesto precoce del contraddittorio tecnico, BONA*, Le indagini digitali tra efficienza investigativa e garanzie difensive, Milano,* GIUFFRÈ*, 2023, p. 112 ss.*

[58] CUROTTI, *Manuale di informatica forense e investigazioni digitali, Torino,* GIAPPICHELLI, *2024, p. 205,* ove si evidenzia come la manipolazione tramite reti generative necessiti di protocolli di congelamento immediato dei flussi di memoria *cache*.

[59]

[?] Sul concetto di copia-specchio e sulla ripetibilità teorica dell'estrazione forense tradizionale, v. la celebre ricostruzione di TARUFFO, *La semplice verità. Il giudice e la costruzione dei fatti, Bari-Roma, Laterza, 2009.* In tal senso, assumono rilievo i principali protocolli operativi in materia di digital forensics, tra cui le Linee guida nazionali per i consulenti tecnici e i periti e lo standard ISO/IEC 27037:2012 relativo

che in questa sede si intende sostenere, di contro, evidenzia la latente fallacia di tale orientamento: l'interazione tra il *software* forense di *detection* e il file multimediale non costituisce un'operazione neutra. L'applicazione di algoritmi diagnostici commerciali i quali operano estraendo e ricombinando i vettori biometrici o i *pixel* del documento può determinare, in assenza di protocolli standardizzati, alterazioni microscopiche non rigenerabili dello stato logico del reperto. Ne consegue che, ogniqualvolta l'accertamento algoritmico implichi una potenziale modificazione, anche solo logica o strutturale, del compendio informativo nativo, il Pubblico ministero ha l'obbligo di attivare le garanzie dell'art. 360 c.p.p., consentendo ai consulenti della difesa di partecipare all'atto sin dal suo avvio.[60] L'estromissione della difesa dall'atto originario di validazione dell'algoritmo si traduce in un pregiudizio irreversibile. Qualora la polizia giudiziaria proceda ad un'estrazione del dato o a un filtraggio *machine-to-machine* in via unilaterale, risulterà successivamente impossibile per l'imputato ricostruire se l'eventuale "falso positivo" sia stato generato da un *bias* nativo del *deepfake* o da un errore metodologico commesso in sede di primo trattamento investigativo. Pertanto, il rispetto delle forme dell'art. 360 c.p.p. assurge a precondizione di utilizzabilità dibattimentale del verdetto tecnologico, configurandosi come l'unico argine processuale capace di preservare la genuinità della fonte e di garantire l'effettività del diritto alla prova scientifica nell'orizzonte della transizione algoritmica.[61]

4. Il giudice custode del metodo scientifico: la metamorfosi della perizia, lo statuto "Cozzini" e il controesame epistemologico dell'algoritmo. Il naturale approdo del modello interpretativo del "doppio binario" e della progressiva ridefinizione del canone dell'oltre ogni ragionevole dubbio si colloca nel momento centrale della formazione della prova: quello in cui il sapere tecno-

all'individuazione, raccolta, acquisizione e conservazione della prova digitale. Tali modelli metodologici vengono talora richiamati anche nell'ambito delle attività di verifica dell'autenticità dei contenuti digitali, pur essendo stati elaborati con riferimento a problematiche differenti.

[60] Questa impostazione estende e adatta i principi espressi dalle Sezioni unite in materia di accertamenti irripetibili sui supporti informatici (Cass. pen., Sez. Un., n. 44320/2021, cit.), evidenziando come l'opacità dell'*AI* introduca una nuova forma di irripetibilità logica dell'atto valutativo: sul punto, FIORIO, *L'accertamento tecnico non ripetibile nell'era della prova algoritmica, in Rivista Italiana di Diritto e proc. pen*, 2024, fasc. 1, p. 240 ss.

[61] TONINI, *Lineamenti di diritto processuale penale applicato alle nuove tecnologie, Milano,* GIUFFRÈ, *2024, p. 315,* il quale conclude rilevando che la violazione delle garanzie partecipative nella prima fase di analisi del dato informatico inficia l'intero percorso epistemologico della decisione giudiziale.

logico viene introdotto nel processo, decodificato e sottoposto al vaglio critico del giudice penale. Se gli strumenti di *deepfake detection* presentano fisiologici margini di errore statistico e, soprattutto, un'intrinseca opacità strutturale derivante dall'impiego di modelli algoritmici complessi, la funzione giurisdizionale non può più esaurirsi nella mera ricezione passiva delle risultanze peritali. Al contrario, il giudice è chiamato a esercitare un penetrante controllo sul metodo scientifico sotteso all'elaborazione tecnica, affinché l'accertamento penale non venga surrettiziamente delegato a soggetti tecnologici privati o a meccanismi computazionali sottratti al contraddittorio processuale.[62] Nel sistema processuale italiano, l'ingresso della prova scientifica atipica trova il proprio riferimento normativo negli artt. 189 e 220 ss. c.p.p. Tuttavia, l'impiego di strumenti di intelligenza artificiale applicati alla *digital forensics* impone una rimeditazione dei tradizionali criteri di valutazione della consulenza tecnica. In questa prospettiva, i criteri elaborati dalla giurisprudenza statunitense nel noto precedente Daubert standard possono assumere rilievo non già quale meccanica trasposizione nel sistema italiano, bensì come paradigma metodologico compatibile con lo statuto epistemologico delineato dalle Sezioni unite sentenza Cozzini, che hanno individuato i requisiti di affidabilità del sapere scientifico utilizzabile nel processo penale.[63] Ne deriva che il giudice, nell'esercizio del proprio potere-dovere di valutazione della prova tecnica, dovrebbe sottoporre gli strumenti algoritmici a un rigoroso scrutinio epistemologico, verificandone anzitutto la controllabilità empirica e la falsificabilità metodologica. Secondo l'impostazione recepita dalla giurisprudenza di legittimità, una teoria può dirsi scientificamente affidabile soltanto se suscettibile di verifica e potenziale confutazione. In tale prospettiva, l'utilizzo di *software* integralmente sottratti a qualsiasi forma di verificabilità esterna perché protetti da segreto industriale assoluto o non replicabili dalla difesa rischia di porsi in tensione con gli artt. 189 e 190 c.p.p., nella misura in cui impedisce alle parti di esercitare un effettivo controllo critico sul procedimento tecnico che ha generato il risultato probatorio. A ciò si aggiunge il necessario

[62] *Per il ruolo del giudice quale garante dell'affidabilità epistemologica della prova scientifica,* TARUFFO, La prova dei fatti giuridici, Milano, GIUFFRÈ, 1992, p. 337 ss.; *nonché* TONINI, La prova penale, Padova, Cedam, ult. ed., p. 215 ss.

[63] *Sentenza Cozzini, Cass. pen., Sez. un., 17 settembre 2010, n. 43786,* Cozzini*, in CED Cassazione. Sul rapporto tra il paradigma Daubert e il sistema italiano,* FERRUA, Il giudizio penale: fatto e valore giuridico*, Torino,* GIAPPICHELLI, *p. 289 ss.*

controllo della comunità scientifica di riferimento. Le metodologie di *deepfake detection* basate, ad esempio, su reti neurali convoluzionali (CNN) o su sistemi di classificazione probabilistica dovrebbero essere supportate da validazioni indipendenti, pubblicazioni scientifiche verificabili e protocolli sperimentali riproducibili. Quanto più il *software* risulti sviluppato in contesti chiusi, privi di revisione esterna o di verifiche indipendenti, tanto maggiore dovrà essere la prudenza del giudice nella valutazione dell'attendibilità dell'elaborato tecnico. Particolare rilievo assume, inoltre, il tema del tasso di errore (*error rate*), che rappresenta uno dei principali parametri di affidabilità scientifica. Il consulente tecnico non può limitarsi a formulare conclusioni assertive circa la presunta autenticità o artificialità del contenuto audiovisivo, ma dovrebbe esplicitare i margini di errore registrati dal sistema su *dataset* comparabili, indicando la percentuale di falsi positivi e falsi negativi rilevati nei test di validazione.[64] In assenza di tali indicazioni, ovvero qualora il margine di errore risulti ignoto o non verificabile, la forza dimostrativa dell'elaborazione algoritmica risulta inevitabilmente ridimensionata, imponendo al giudice una valutazione particolarmente rigorosa ai fini del rispetto della regola dell'oltre ogni ragionevole dubbio. Parimenti centrale appare il tema dell'indipendenza dell'esperto e della neutralità metodologica del sistema impiegato. Nel settore dell'intelligenza artificiale forense, in cui gli strumenti sono frequentemente sviluppati da soggetti privati portatori di interessi economici o istituzionali, il giudice dovrebbe verificare non soltanto la competenza tecnica del consulente, ma anche la qualità dei dati di addestramento utilizzati dal modello, al fine di escludere la presenza di *bias* sistematici suscettibili di alterare l'esito dell'analisi. L'intero protocollo di acquisizione, conservazione ed elaborazione del dato digitale dovrebbe inoltre conformarsi agli standard internazionali della *digital forensics,* inclusi quelli previsti dalla norma ISO/IEC 27037, con particolare attenzione alla tracciabilità delle operazioni compiute e alla preservazione dell'integrità del reperto informatico. Sotto il profilo del diritto di difesa, l'impiego della prova algoritmica impone una valorizzazione del contraddittorio tecnico quale proiezione degli artt. 24 e 111 Cost. La verificabilità del metodo utilizzato non costituisce una mera esigenza scientifica, ma rappresenta il presupposto indispensabile affinché la difesa possa contestare efficacemente l'attendibilità del risultato prodotto dall'algoritmo. In tale prospet-

[64] *National Institute of Standards and Technology (NIST),* AI-generated Content and Deepfake Detection: Technical Challenges and Limitations.

tiva, l'esame incrociato del consulente tecnico *ex* art. 501 c.p.p. tende progressivamente a trasformarsi in un controllo critico sul funzionamento del sistema impiegato, consentendo alla difesa di interrogare non soltanto il perito, ma anche la solidità epistemologica del procedimento computazionale sottostante. Ciò assume particolare rilievo rispetto al rischio di automation bias, ossia alla tendenza dell'interprete umano ad attribuire all'elaborazione tecnologica una presunzione di neutralità e infallibilità. Proprio per tale ragione, quanto minore sia il grado di trasparenza del sistema algoritmico, tanto maggiore dovrà essere l'onere motivazionale gravante sul giudice e tanto più rigorosa dovrà risultare la ricerca di elementi di riscontro esterni idonei a corroborare l'esito tecnico. In questa prospettiva, la tutela del cittadino nell'ecosistema digitale non sembra potersi realizzare esclusivamente mediante l'inasprimento sanzionatorio delle condotte di manipolazione audiovisiva, ma richiede anche un adeguamento delle garanzie processuali relative all'utilizzo della prova tecnologica. Più che introdurre automatismi di esclusione generalizzata, appare necessario rafforzare normativamente gli obblighi di trasparenza, tracciabilità e verificabilità delle metodologie algoritmiche utilizzate nel procedimento penale,[65] così da preservare la centralità del giudice quale *peritus peritorum* e garantire l'effettività del principio di parità delle armi. In tale scenario, assume crescente importanza anche la predisposizione di sistemi di certificazione dell'integrità del dato digitale, fondati su protocolli di *timestamping* crittografico e su tecniche di attestazione immutabile della catena di custodia informatica.[66] Là dove il reperto digitale non presenti adeguate garanzie di integrità e tracciabilità, il giudice dovrà valutarne con particolare cautela l'affidabilità probatoria. Sotto questo profilo, l'impiego di metodologie algoritmiche di *detection* prive di standardizzazione normativa introduce un elemento di atipicità metodologica *ex* art. 189 c.p.p., imponendo al giudicante di verificare in concreto se tale strumento di accertamento conservi requisiti minimi di attendibilità scientifica compatibili con il giusto processo.

[65] *European Union, Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (AI Act).*

[66] *European Union Agency for Cybersecurity (ENISA),* Guidelines on Secure AI Systems.

5. I rimedi impugnatori e lo scrutinio di legittimità. Le problematiche derivanti dall'impiego della prova algoritmica[67] non si esauriscono nella fase dibattimentale, ma si riflettono inevitabilmente anche sul sistema delle impugnazioni e, in particolare, sul controllo esercitato dalla Corte di cassazione ai sensi dell'art. 606, co. 1, lett. e) c.p.p. La giurisprudenza di legittimità ha da tempo chiarito che il sindacato sulla prova scientifica non concerne la preferibilità astratta di una determinata teoria, bensì la correttezza metodologica del percorso argomentativo seguito dal giudice di merito.[68] Tuttavia, l'ingresso degli strumenti di intelligenza artificiale nel procedimento probatorio impone un rafforzamento dell'obbligo motivazionale, specialmente nei casi in cui la decisione si fondi in misura significativa sul responso di *software* di *deepfake detection* o di analisi automatizzata del contenuto audiovisivo.[69] Il vizio di motivazione può assumere, in questo contesto, almeno due distinte configurazioni. Una prima ipotesi riguarda il travisamento della prova probabilistica, che si verifica quando il giudice attribuisca valore sostanzialmente assoluto a un risultato espresso in termini meramente statistici. Qualora, ad esempio, il sistema indichi una probabilità di alterazione pari al 92%, il giudice è tenuto a spiegare attraverso quali ulteriori elementi di riscontro ritenga superata la residua area di incertezza compatibile con il ragionevole dubbio. Una seconda ipotesi concerne il travisamento del fatto digitale, ravvisabile laddove la decisione assuma come storicamente autentica una rappresentazione audiovisiva che le risultanze tecniche o le deduzioni difensive indicavano come potenzialmente manipolata o artificialmente generata. In tali casi, l'omessa valutazione critica degli elementi tecnici contrari può tradursi in un vizio motivazionale idoneo a compromettere la tenuta logica dell'intero impianto decisionale.[70] Ne consegue che la parte ricorrente, nel proporre ricorso per cassazione, non potrà limitarsi a contestazioni generiche circa l'inattendibilità del pro-

[67] Ai fini della presente trattazione, l'espressione "prova algoritmica" non si riferisce ai noti modelli predittivi di orientamento della decisione penale o di valutazione del rischio di recidiva, bensì all'impiego di sistemi di intelligenza artificiale sia nella fase di genesi dell'evidenza multimediale manipolata (*deepfake*), sia nei successivi protocolli tecnologici di *detection* scientifico-forense.

[68] *Cass. pen., sez. IV, 13 dicembre 2017, n. 11197,* in tema di controllo di legittimità sulla motivazione relativa alla prova scientifica.

[69] Sul dovere di motivazione rafforzata rispetto alla prova tecnico-scientifica, GIOSTRA, *Prima lezione sulla giustizia penale, Bari-Roma, Laterza, p. 143 ss.*

[70] *I*n tema di travisamento della prova e omissione valutativa di elementi decisivi*, v. Cass. pen., Sez. VI, 14 febbraio 2013, n. 8700.*

gramma utilizzato, ma dovrà allegare gli elementi tecnici specifici quali protocolli di validazione, *dataset* di controllo o passaggi del controesame del consulente idonei a dimostrare la concreta frattura metodologica verificatasi nel giudizio di merito. Particolare attenzione merita altresì il caso della prova dichiarativa formatasi attraverso la percezione mediata di contenuti audiovisivi successivamente rivelatisi manipolati. Anche in tale ipotesi, il controllo di legittimità dovrà investire la coerenza logica della motivazione che abbia ritenuto attendibile la deposizione senza verificare criticamente l'affidabilità del supporto digitale da cui essa traeva origine. In definitiva, il sindacato della Corte di cassazione rappresenta uno strumento essenziale per impedire che la decisione penale si trasformi in una mera ratifica dell'esito computazionale prodotto dalla macchina. Il controllo di legittimità non investe il *software* di *detection* in sé considerato, ma la razionalità del percorso argomentativo mediante il quale il giudice ha ritenuto di attribuire efficacia probatoria al dato digitale cosi verificato. Solo mantenendo centrale il controllo umano sulla motivazione della decisione è possibile ricondurre il progresso tecnologico entro i confini costituzionali del giusto processo.[71] L'esigenza di sottoporre la prova algoritmica a rigorosi criteri di verificabilità non implica, tuttavia, l'impossibilità di utilizzare strumenti tecnologici nel processo penale. L'ordinamento dispone già di categorie dogmatiche idonee a governare il fenomeno, purché il giudice eviti di attribuire al dato digitale un valore autosufficiente e dirimente. In tale prospettiva, qualora l'accertamento tecnico restituisca indici probabilistici di alterazione o margini di dubbio scientifico, il responso algoritmico dovrebbe essere ricondotto nell'ambito della prova indiziaria disciplinata dall'art. 192, comma 2, c.p.p. L'esito probabilistico fornito dal *software* di *detection* non appare infatti assimilabile a una prova rappresentativa diretta, ma costituisce un elemento inferenziale che necessita di essere corroborato da ulteriori riscontri esterni caratterizzati dai requisiti di gravità, precisione e concordanza.[72] La verifica dell'autenticità del contenuto audiovisivo non può pertanto esaurirsi nell'analisi tecnica del *file*, ma richiede la sua contestualizzazione all'interno dell'intero compendio probatorio. Il giudice dovrà valutare se il contenuto digitale trovi conferma in ulteriori elementi

[71] Sul controllo della Cassazione quale presidio di razionalità della motivazione, SCALFATI, La motivazione della sentenza penale, *Torino,* GIAPPICHELLI, *p. 201 ss.*

[72] Per la natura e i requisiti della prova indiziaria *ex* art. 192 c.p.p., SPANGHER, *Procedura penale, Milano, ult. ed., p. 489 ss.*

acquisiti al processo, quali dichiarazioni testimoniali, dati di geolocalizzazione, tabulati telefonici, accertamenti tecnici o ulteriori evidenze documentali coerenti con la ricostruzione accusatoria. Laddove il responso algoritmico rimanga isolato e privo di riscontri indipendenti, la sua capacità dimostrativa risulterà inevitabilmente insufficiente ai fini dell'affermazione di responsabilità penale. Viceversa, quando l'elemento tecnico si inserisca coerentemente in un quadro indiziario convergente e logicamente strutturato, la regola dell'oltre ogni ragionevole dubbio potrà ritenersi rispettata non già per una presunta infallibilità della macchina, ma per la complessiva tenuta razionale del compendio probatorio valutato dal giudice umano.

6. *Considerazioni conclusive: per un'epistemologia costituzionale della resistenza algoritmica.* Il viaggio intrapreso lungo i tornanti della dematerializzazione della sanzione e della prova, sotto lo spettro immanente dell'*infocalypse* audiovisiva, conduce a un approdo che non è meramente tecnico, ma profondamente costituzionale ed assiologico. La scomposizione dogmatica del fenomeno dei *deepfake* e l'analisi dei correlati strumenti di *detection* hanno disvelato come la transizione digitale non rappresenti una semplice evoluzione delle modalità di commissione o di accertamento dei reati, bensì una radicale sfida allo statuto epistemologico della giurisdizione penale. La traiettoria evolutiva sin qui tracciata ha evidenziato il rischio concreto di una progressiva capitolazione del processo penale di fronte all'«illusione del vero algoritmico». L'ordinamento, dinanzi alla vertiginosa fluidità del dato cross-mediale artificiale, si trova sospeso tra due derive simmetriche e altrettanto letali: da un lato, l'insorgenza di uno scetticismo giudiziale paralizzante, capace di azzerare la funzione cognitiva del processo penale; dall'altro, una supina ed acritica delega del giudizio di colpevolezza a stringhe di codice privato, governata dall'insidioso fenomeno dell'*automation bias* e dalla fallace percezione di un'infallibilità della macchina. Rispetto a tali derive, le risposte fornite dalla politica penale contemporanea pervicacemente arroccata su un modello di ipertrofia sanzionatoria e sulla sterile creazione di nuove fattispecie incriminatrici si rivelano declamatorie e inefficaci se disgiunte da una vigorosa riscrittura delle regole del rito. La crisi della legalità sostanziale, che affoga nell'indeterminatezza di condotte dematerializzate e difficilmente riconducibili ai paradigmi classici dell'offensività, trova la sua reale linea di resistenza non già nell'inasprimento delle pene edittali, bensì nella blindatura dello statuto delle garanzie proces-

suali. La tutela del cittadino nell'era degli artefatti digitali si realizza governando rigorosamente il momento genetico, acquisitivo e valutativo della prova tecnologica, poiché è la tenuta del metodo processuale a fungere da ultimo baluardo di legalità laddove la norma sostanziale evapora nella complessità della rete. La vera linea di difesa del cittadino nell'era degli artefatti digitali non si traccia inasprendo le pene edittali, ma blindando lo statuto delle garanzie processuali nel momento genetico, acquisitivo e valutativo della prova tecnologica. La proposta dogmatica formulata nel presente contributo, imperniata sul modello del "doppio binario" interpretativo, mira precisamente a restituire all'organo giudicante l'effettivo governo del sapere scientifico. Configurare la totale trasparenza dell'algoritmo (ostensibilità dei *training data* e delle metriche di errore) quale precondizione inderogabile di ammissibilità della prova *ex* artt. 189 e 190 c.p.p. significa riaffermare l'invalicabilità del principio del contraddittorio e della parità delle armi (artt. 24 e 111 Cost.). Al contempo, interpretare il tasso di errore statistico intrinseco della macchina come oggettiva e formale espressione del "ragionevole dubbio" *ex* art. 533, co. 1 c.p.p. impedisce l'utilizzazione del solo dato computazionale ai fini della pronuncia di condanna, costringendo il giudice a rintracciare la storicità del fatto nel perimetro reale di un quadro indiziario coordinato e storicizzato *ex* art. 192, co. 3, c.p.p. Solo attraverso questa rigorosa sottomissione del progresso tecnologico alle regole della grammatica codicistica e costituzionale, il processo penale può conservare la propria legittimazione democratica. Il Giudice, ergendosi a strenuo custode del metodo scientifico attraverso lo statuto *Cozzini-Daubert* e governando la dialettica del controesame epistemologico della macchina, riafferma il proprio ruolo non surrogabile di *peritus peritorum*. In caso contrario, l'ordinamento abdicherebbe al proprio mandato, convertendo l'accertamento della responsabilità penale in una supina ratifica di un verdetto computazionale transnazionale. La tenuta del canone del giusto processo nell'era dei *deepfake* risiede, in ultima analisi, nella capacità del diritto di esercitare una consapevole ed intransigente resistenza algoritmica, a presidio intramontabile della libertà personale e della dignità dell'uomo. In conclusione, il passaggio dalla giustizia documentale alla giustizia algoritmica non deve tradursi in una delega della funzione giurisdizionale. Il canone dell'«oltre ogni ragionevole dubbio» (ex art. 533 c.p.p.) oggi si colora di una nuova sfumatura: il dubbio sulla genuinità ontologica del supporto. Se la macchina può generare una verità alternativa indistinguibile dal vero, il libero convincimento del giu-

dice deve tornare a farsi baluardo antropocentrico. La sfida per il diritto penale contemporaneo è dunque quella di presidiare i confini del "giusto processo", garantendo che la sentenza resti un atto di responsabilità umana, unico argine contro l'impersonalità dell'errore algoritmico.