VERSO IL SUPERAMENTO DEL "*LEGAL RISK*" EUROPEO: INTELLIGENZA ARTIFICIALE E APPROCCIO PROPORZIONALE AL RISCHIO

di Maria Carla Canato
(*Dottoranda di ricerca in diritto penale, Università degli Studi di Padova*)

1. L'*AI Act*[1] *(AIA)* costituisce, nel panorama normativo europeo, il primo tentativo di fornire un quadro giuridico armonizzato per disciplinare, estensivamente, la progettazione, lo sviluppo, e l'utilizzo dei sistemi di intelligenza artificiale. In quanto teso a promuovere lo sviluppo, l'uso e l'adozione dell'IA nel mercato interno, e – al tempo stesso – a garantire un elevato livello di protezione degli interessi pubblici, quali la salute, la sicurezza e la protezione dei diritti fondamentali, come riconosciuti e tutelati dal diritto dell'Unione[2], potrebbe essere definito, in senso ampio, alla stregua di una "Costituzione europea" in tema di IA"[3].

Anche in questo particolare ambito, si sta quindi realizzando nel contesto comunitario una progressiva opera di "ridefinizione" – non solo giuridica ma anche

[1] Risoluzione legislativa del Parlamento europeo del 13.3.2024 sulla proposta di regolamento del Parlamento europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale) e modifica alcuni atti legislativi dell'Unione (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)). Entrerà in vigore (180 Considerando, 113 Articoli, 13 Allegati) 20 giorni dopo la pubblicazione nella Gazzetta ufficiale UE e inizierà ad applicarsi 24 mesi dopo l'entrata in vigore salvo per quanto riguarda: i divieti relativi a pratiche vietate, che si applicheranno a partire da sei mesi dopo l'entrata in vigore; i codici di buone pratiche (9 mesi dopo); le norme sui sistemi di IA per finalità generali, compresa la *governance* (12 mesi dopo); gli obblighi per i sistemi ad alto rischio (36 mesi dopo).

[2] Tali sono gli scopi espressamente esplicitati nel Considerando 5 della Proposta di Regolamento 2021/0106 (COD).

[3] C. Novelli, *L'Artificial Intelligence Act Europeo: alcune questioni di implementazione*, in *federalismi.it*, 2, 2024, 95.

culturale – cui viene riconnessa una "globalizzazione della tutela" dei diritti fondamentali[4]. Con l'evoluzione tecnologica, infatti, i diritti umani, come aveva evidenziato la stessa Corte di Strasburgo già dalla fine degli anni Settanta[5], sono da intendersi come «*a living instrument which… must be interpreted in the light of present-day conditions*», e stanno egualmente subendo una significativa fase di trasformazione[6].

Di conseguenza, in questo mutato assetto di principi di rilevanza costituzionale e alla luce del dibattito che, da lungo tempo, ha interessato la dottrina penalistica nazionale e sovranazionale, è innegabile che lo studioso del diritto debba porsi nuovi interrogativi in merito ai profili di responsabilità – "diretta" della macchina e "indiretta", o vicaria, dell'uomo[7] – derivanti dall'utilizzo di queste tecnologie innovative[8], la cui evoluzione non impone necessariamente un significativo e costante

[4] Ciò emerge, oltretutto, nell'ambito del Consiglio d'Europa, dalla 'Carta etica europea per l'uso dell'intelligenza artificiale nei sistemi di giustizia penale e nei relativi ambienti'. Il documento è stato stilato dalla CEPEJ, Commissione europea per l'efficacia della giustizia, istituita nel 2002 per iniziativa del Comitato dei Ministri del Consiglio d'Europa, con lo scopo di monitorare e misurare la qualità dei sistemi giudiziari dei Paesi membri. Numerosi sono i principi affermati dalla Carta con riferimento all'uso dei sistemi di IA. In particolare, si richiamano il rispetto dei diritti fondamentali, il principio di non discriminazione, il principio di qualità e sicurezza, i principi di trasparenza, imparzialità e *fairness*, oltre al controllo da parte dell'utente.
Per quanto attiene alle problematiche riconnesse all'uso di sistemi di IA nel contesto della giustizia penale, in aggiunta, lo studio annesso alla Carta prospetta l'inadeguatezza del termine 'giustizia predittiva', atteso che molti dei dati impiegati per la valutazione di rischio di comportamenti violenti, diffusamente utilizzati nelle corti americane, determinano un'accentuazione delle discriminazioni già esistenti. Inoltre, l'impiego di specifici *softwares* predittivi, anche ai fini della valutazione del *risk assessment,* deve essere subordinato al noto *Daubert test*, il quale impone che ogni metodo sia sottoposto a riproduzione e falsificazione, attraverso la pubblicazione scientifica e la *peer review*. Per una disamina più completa del tema, si rimanda a S. 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*, 2018.

[5] Vedasi la pronuncia *Tyrer v. the United Kingdom* App no 5856/72 (ECtHR, 25 April 1978), para 31.

[6] A. Ortalda-P. De Hert, *Artificial Human Rights Impact Assessment*, in *Artificial Intelligence and Human Rights*, a cura di A. Quintavalla e J. Temperman, Oxford 2023, 532. Gli autori affermano infatti che «*Not only technology changes. Human rights are equally undergoing a period of change*».

[77] Sui due modelli di imputazione (*robot* come "strumento" e come "soggetto" del reato), per tutti, vedasi S. Riondato, *Robot: talune implicazioni di diritto penale*, in *Tecnodiritto. Temi e problemi di informatica e robotica giuridica*, a cura di P. Moro, C. Sarra, Milano 2017, 85 ss.

[8] Vedasi, *ex multis*, *Traditional Criminal Law Categories and AI: Crisis or Palingenesis?*, a cura di L. Picotti- B. Panattoni, Antwerpen 2023; F. Basile, *Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine*, in *www.dirittopenaleuomo.org*, 29.9.2019, 24 ss.; M.B. Magro, *Decisione umana e decisione robotica. Un'ipotesi di responsabilità da procreazione robotica*, in *www.lalegislazionepenale.eu*, 10.5.2020, 4 ss.; S. Gleß-T. Weigend, *Intelligente Agenten und das Strafrecht*, in *ZStW* 2014, 561 ss.; G. Hallevy, *Liability for Crimes Involving Artificial Intelligence Systems*, Berlin-New York 2014, 185 ss.; E. Hilgendorf, *Können Roboter schuldhaft handeln?*, in *Jenseits von Mensch und Maschine. Ethische und rechtliche Fragen zum Umgang mit Robotern, Künstlicher Intelligenz und Cyborgs*, a cura di S. Beck, Baden-Baden 2012, 119 ss.; M. Simmler-N. Markwalder, *Roboter in der Verantwortung? Zur Neuauflage der Debatte um den funktionalen Schuldbegriss,* in *ZStW* 2017, 20 ss.; S. Ziemann,

controllo umano. Infatti, molte macchine intelligenti sono oggi in grado di operare[9], senza alcuna supervisione umana [10], assumendo, tramite i meccanismi di apprendimento propri del *machine learning*, caratteristiche di imprevedibilità [11]. Il soggetto artificiale "ricorda" il passato, apprende dal proprio "vissuto", modificando e adattando conseguentemente il proprio comportamento[12].

Addirittura, in molti casi l'IA può acquisire informazioni anche l'esperienza dei suoi

Wesen, seid's gewesen? Zur Diskussion über ein Strafrecht für Maschinen, in *Robotik und Gesetzgebung,* a cura di E. Hilgendorf e J.P. Günther, Baden-Baden 2013, 183 ss.; F. Basile, *Diritto penale e intelligenza artificiale*, in *GI* 2019, 69 ss.; U. Pagallo-S. Quattrocolo, *The impact of AI on criminal law, and its twofold procedures*, in *Research Handbook on the Law of Artificial Intelligence*, a cura di W. Barfield e U. Pagallo, Northampton 2018, 385 ss.; D. Lima, *Could AI Agents Be Held Criminally Liable? Artificial Intelligence and the Challenges for Criminal Law*, in *South Carolina Law Review* 2018, 677 ss.; T. King, N. Aggarwal, M. Taddeo, L. Floridi, *Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solution*s, in *Science and Engineering Ethics* 2019, *passim*; P. Asaro, *A body to Kick, but Still No Soul to Damn: Legal Perspectives on Robotics*, in *Robot Ethics*, a cura di P. Lin, K. Abney e G.A. Bekey, Cambridge 2012, 169 ss.; S. Riondato, *Robot: talune implicazioni di diritto penale*, cit., 87 ss.; T.C. King, N. Aggarwal, M. Taddeo, L. Floridi, *Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions*, in *Science and Engineering Ethics*, anticipato *online* il 14.2.2019.

[9] In tal senso, anche la Risoluzione del Parlamento europeo del 16 febbraio 2017, recante raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica (2015/2103(INL)) - pur non affrontando i profili attinenti alla materia penalistica pone l'accento sul principio di trasparenza, nonché sui profili di autonomia e di imprevedibilità dei sistemi di AI: «AA. considerando che l'autonomia di un robot può essere definita come la capacità di prendere decisioni e metterle in atto nel mondo esterno, indipendentemente da un controllo o un'influenza esterna; che tale autonomia è di natura puramente tecnologica e il suo livello dipende dal grado di complessità con cui è stata progettata l'interazione di un robot con l'ambiente; AB. considerando che più i robot sono autonomi, meno possono essere considerati come meri strumenti nelle mani di altri attori (quali il fabbricante, l'operatore, il proprietario, l'utilizzatore, ecc.); che ciò, a sua volta, pone il quesito se le regole ordinarie in materia di responsabilità siano sufficienti o se ciò renda necessari nuovi principi e regole volte a chiarire la responsabilità legale dei vari attori per azioni e omissioni imputabili ai robot, qualora le cause non possano essere ricondotte a un soggetto umano specifico, e se le azioni o le omissioni legate ai robot che hanno causato danni avrebbero potuto essere evitate; AC. considerando che, in ultima analisi, l'autonomia dei robot solleva la questione della loro natura alla luce delle categorie giuridiche esistenti e dell'eventuale necessità di creare una nuova categoria con caratteristiche specifiche e implicazioni proprie [...]; IA. considerando che [...] l'attuale quadro giuridico non sarebbe sufficiente a coprire i danni causati dalla nuova generazione di robot, in quanto questi possono essere dotati di capacità di adattamento e di apprendimento che implicano un certo grado di imprevedibilità nel loro comportamento, dato che imparerebbero in modo autonomo, in base alle esperienze diversificate di ciascuno, e interagirebbero con l'ambiente in modo unico e imprevedibile».

[10] I. Salvadori, *Agenti artificiali, opacità tecnologica e distribuzione della responsabilità penale*, in *RIDPP* 2021, 90 ss.

[11] Merita di essere richiamata, in tal senso, la distinzione tra algoritmi "impliciti" ed "espliciti" delineata da A. Jean, *Les algorithmes. Font-ils la loi?*, Paris 2021, 32 ss.

[12] E. Palmerini, voce *Robotica*, in *Enciclopedia di bioetica e scienza giuridica*, diretta da E. Sgreccia-A. Tarantino, X, Napoli 2016, 1106. Sul *machine learning* in ambito giuridico, per tutti, v. H. Surden, *Machine Learning and Law*, in *Washington Law Review* 2014, 87 ss.; D.R. Desai, *Exploration and Exploitation. An Essay on (Machine) Learning, Algorithms, and Information Provision*, in *Loyola University Chicago Law Journal* 2015, 541 ss.; J. Stilgoe, *Machine learning, social learning and the governance of self-driving cars*, in *Social Studies of Science* 2018, 29 ss.; A. Bertolini, *Robots and liability. Justifying a change in perspective*, in *Rethinking Responsibility in Science and Technology*, a cura di F. Battaglia, N. Mukerji e J. Nida-Rümelin, Pisa 2014, 155 ss.

"simili", mediante il ricorso alle tecnologie di *cloud computing*[13]. Ciò consente di "sommare" le "esperienze di vita" di una moltitudine di macchine intelligenti, sottoposte agli scenari più diversi. Non a caso, alcune applicazioni sono spesso utilizzate per incrementare esponenzialmente la rapidità di apprendimento di tali soggetti robotici, con conseguente riduzione del grado di controllo da parte dell'uomo[14]. Il comportamento delle macchine in tal modo diviene, di riflesso, *ex ante* in parte imprevedibile, da un punto di vista non solo soggettivo, ma anche oggettivo-tecnologico[15], nell'ambito dei c.d. *black box algorithms*[16].

L'elevato livello di "autonomia" del sistema può peraltro, in taluni casi, essere idoneo a determinare danni riconducibili e connessi a processi di elaborazione "imperscrutabili"[17]. Al punto che recentemente una parte della dottrina ha ipotizzato la necessità di teorizzare una responsabilità penale "diretta" [18] delle macchine,

[13] La tecnologia in parola è molto studiata, in ragione dei suoi collegamenti con la filosofia del diritto internazionale e il diritto internazionale umanitario che con il diritto penale comune. Ai fini di questo studio necessariamente sintetico, ci si limita a richiamare alcuni lavori fondamentali di P. Asaro, *The labor of surveillance and bureaucratized killing: new subjectivities of military drone opera-tors*, in *Social Semiotics* 2013, 1 ss.; Id., *On banning autonomous weapon systems: human rights, au-tomation, and the dehumanization of lethal decision-making*, in *International Review of the Red Cross* 2012, 687 ss. Più in generale, sul tema dei droni, vedasi anche *Drones and Unmanned Aerial Systems*, a cura di B. Završnik, Berlino 2016.

[14] A. Cappellini, *Machina delinquere non potest? Brevi appunti su intelligenza artificiale e responsabilità penale*, in *Criminalia* 2019, 6 ss.

[15] Per tutti, U. Pagallo, *The Laws of Robots. Crimes, Contracts and Torts*, Berlino 2013, 47; S. Beck, *Google Cars, Software Agents, Autonomous Weapons Systems – New Challanges for Criminal Law?*, in *Law, Computer Science* 2017, 243. Rispetto alle *autonomous cars*, cfr. H. Surden-M.A. Williams, *Technological Opacity, Predictability, and Self-Driving Cars*, in *Cardozo Law Review* 2016, 157 ss.

[16] Con tale espressione, si fa riferimento alla situazione in cui tra i dati di *input* e i comportamenti tenuti come *output* vi sia un'opacità, un vuoto di comprensione da parte dell'osservatore umano esterno, così che le condotte di queste IA siano gravate, in un'ottica a priori, da un ineliminabile margine di imponderabilità. Sul punto, vedasi S. Doncieux-J.B. Mouret, *Beyond black-box optimization: a review of selective pressures for evolutionary robotics*, in *Evolutionary Intelligence* 2014, 71 ss.

[17] M.E. Florio, *Il dibattito sulla responsabilità penale diretta delle IA: "molto rumore per nulla"?*, in *www.sistemapenale.it*, 2024, 2, 6 ss.

[18] La dottrina di riferimento in materia, cui si rinvia per ulteriori approfondimenti, è primariamente riconducibile alle teorie del giurista israeliano Gabriel Hallevy. Vedasi, in particolare, G. Hallevy, *The Criminal Liability of Artificial Intelligence Entities – from Science Fiction to Legal Social Contro*l, in *Akron Intellectual Property Journal* 2010, 171 ss.; Id., *"I, Robot – I, Criminal"–– When Science Fiction Becomes Reality: Legal Liability of AI Robots committing Criminal Offences*, in *Syracuse Science & Technology Law Reporter* 2010, 1 ss.; Id., *Virtual Criminal Responsibility*, in *Original Law Review* 2010, 6 ss.; Id., *Dangerous Robots – Artificial Intelligence vs. Human Intelligence*, in SSRN, 21.2.2018. Solo apparentemente più settoriali, v. anche Id., *Unmanned Vehicles: Subordination to Criminal Law under the Modern Concept of Criminal Liability*, in *Journal of Law, Information and Science* 2012, 200 ss.; Id., *AI v. IP. Criminal Liability for IP Offences of AI Entities*, in SSRN, 18.11.2015. L'autore ha dedicato due monografie al tema. La prima, edita negli USA (Boston) nel 2013, con un formato a metà tra lo scritto scientifico e quello destinato al grande pubblico, reca il titolo *When Robots Kill. Artificial Intelligence under Criminal Law*. La seconda, puramente accademica, edita da Springer (Dordrecht) nel 2015 con il titolo

principalmente sollecitata da quegli "automi" che si affidano ad algoritmi in tutto o in parte "impliciti".

Queste circostanze, ormai particolarmente diffuse, fanno cioè entrare in crisi l'approccio penalistico tradizionale alle *res mechanicae*. L'impostazione dogmatica del modello imputativo vicario non appare, in altri termini, idonea a ricondurre la penale responsabilità ai soggetti programmatori o utilizzatori degli strumenti di IA, atteso che, diversamente opinando, si verrebbero a realizzare forme vietate di responsabilità oggettiva.

Tali criticità, nondimeno, riguardano solo marginalmente le ipotesi dei reati a matrice dolosa, ove la macchina rappresenta soltanto lo strumento attraverso cui viene perpetrata la condotta commessa dall'uomo.

In questi casi, la presenza della tensione verso il risultato – che è la volontà, tipica del dolo – fa prescindere in larga misura dal concreto svolgimento del percorso causale attivato, purché l'evento voluto si sia poi effettivamente verificato. Ciò accade sempreché i tratti assunti in concreto dall'evento rientrino nei limiti della sua descrizione astratta, già presente *ex ante* nella mente dell'attore umano. In tale ipotesi, infatti, una qualunque deviazione imprevedibile del comportamento della macchina si risolverebbe in una mera *aberratio causae*, la quale, come noto, tradizionalmente, non fa venire meno il dolo rispetto al fatto[19].

Il dolo, peraltro, rimane presente anche, all'opposto, nella circostanza in cui la deviazione, a priori imprevedibile, nel comportamento del soggetto artificiale produca la mancata verificazione dell'evento pur voluto dall'utilizzatore umano. Peraltro, non appare nemmeno esclusa, in quest'ultima ipotesi, la configurabilità del delitto tentato, sempre che ricorrano i requisiti dell'idoneità e inequivocità degli atti[20].

Liability for Crimes Involving Artificial Intelligence Systems, affronta e teorizza, sotto ogni profilo, il tema generale della responsabilità penale dei soggetti artificiali (cfr. *Preface*, V).

[19] F. Palazzo, *Corso di diritto penale*[7], Torino 2018, 290-291.

[20] Sul punto, si rimanda a S. Beck, *Google Cars, Software Agents, Autonomous Weapons Systems – New Challanges for Criminal Law*, cit., 236. Le criticità emergono, specificatamente, rispetto all'imputazione di tutti quei comportamenti imprevedibili e non voluti dall'utilizzatore umano, rispetto ai quali non appare possibile l'applicazione delle categorie tradizionali del dolo e della colpa. Quanto poi alla preterintenzione, nondimeno, ciò vale soltanto nel caso in cui si ritenga di aderire all'interpretazione costituzionalmente conforme dell'istituto, la quale esclude la rimproverabilità a titolo di responsabilità oggettiva, imponendo - anche in tale contesto - l'individuazione di una colpa in concreto rispetto all'evento ulteriore non voluto (in tal senso, cfr. F. Mantovani, *Diritto penale. Parte generale*[8], Padova, 2013, 365-366). Quanto al profilo problematico dell'*aberratio delicti* nel contesto delle condotte imprevedibili perpetrate dal sistema di IA, cfr. U. Pagallo, *What Robots Want: Autonomous Machines Codes and New Frontiers of Legal Responsibility*, in M. Hildebrandt, *Human Law and Computer Law: Comparative Perspectives*, Berlino 2013, 51-52.

Le criticità cui si è fatto cenno, piuttosto, emergeranno con forza nel settore della colpa, in cui la commissione di reati tramite *AI tools* si prevede possa avere in futuro un rilievo numerico assolutamente preponderante[21].

Le figure umane che possono essere considerate ai fini dell'imputazione, a questo titolo, sono plurime.

Anzitutto gli utilizzatori, purché resti in capo loro un qualche potere impeditivo dell'evento[22]. Al manufattore ben potranno essere contestati errori di montaggio; così come al programmatore saranno imputabili mancanze classificabili come veri e propri errori di programmazione, ovvero nei casi in cui non è stato previsto ciò che si poteva – e doveva – prevedere. In tale ipotesi, ci si muove nel contesto della responsabilità per danno da prodotto[23].

Quando il comportamento del sistema di IA è privo dei connotati di prevedibilità ed evitabilità, tuttavia, le "risposte" del diritto penale tendono a vacillare.

Il principio di colpevolezza, come inteso a livello sovranazionale, dovrebbe infatti ricomprendere non solo il divieto di responsabilità per fatto altrui, ma anche la prevedibilità delle conseguenze sanzionatorie frutto di una determinata condotta[24].

In questo caso, in particolare, oltre ai problemi che potrebbero sorgere dal punto di vista della determinazione del nesso di causalità[25], l'imputazione penale a titolo di

---

[21] S. Gless, E. Silverman, T. Weigend, *If robots cause harm, who is to blame? Self-driving cars and criminal liability*, in *New Criminal Law Review,* 2016, 425 ss.

[22] E. Hilgendord, *Automated Driving and the Law*, in *Robotics, Autonomics and the Law*, a cura di E. Hilgendorf-U. Seidel, Baden-Baden 2017, 181-182.

[23] Si veda, *ex multis*, l'ampia monografia di C. Piergallini, *Danno da prodotto e responsabilità penale. Profili dommatici e politico-criminali*, Milano 2004, oppure più sinteticamente Id., *La responsabilità del produttore: una nuova frontiera del diritto penale?*, in *DPP* 2007, 1125 ss. Rapportano la responsabilità vicaria colposa del r*obot* all'ambito della responsabilità per danno da prodotto S. Gless, E. Silverman, T. Weigend, *If robots cause harm, who is to blame?*, cit., 426 ss. Da ultimo, in materia di *self driving cars*, vedasi il contributo di R. Compostella, *Auto a guida autonoma e diritto penale. Profili di responsabilità individuale e collettiva*, Trento 2024, 102 ss.

[24] *Ex multis*, cfr. C. eur., 14.4.2015, *Contrada c. Italia*. Più di recente, v. C. eur., GC, 23.2.2017, *De Tommaso c. Italia.* Quanto alla nozione di legge penale adottata dalla Corte EDU, deve anzitutto evidenziarsi che essa fa riferimento alla materia penale di tipo sostanziale, individuata – al fine di eludere la c.d. frode delle etichette – con riferimento ai c.d. *Engels criteria*. Tale assunto è maggiormente riferibile al più ampio concetto di diritto che non a quello di legge. In questa (ri)lettura della 'legge' e, più in generale, della legalità penale come tradizionalmente intesa nei sistemi di *civil law*, deve considerarsi che la Corte EDU ha elaborato, quale requisito minimo di "qualità" della legge, la ragionevole conoscibilità delle norme e la ragionevole prevedibilità dell'applicazione delle stesse. Conseguentemente, i requisiti di conoscibilità e prevedibilità devono essere letti con riferimento al diritto vivente non potendosi più restare legati esclusivamente al diritto positivo. L'art. 7 CEDU è quindi considerato *an essential element of the rule of law*, assurgendo a principio inderogabile.

[25] Per un primo inquadramento delle questioni che potrebbero porsi su tale fronte, cfr., nella dottrina italiana, D. Piva, *Machina discere, (deinde) delinquere et puniri potest*, in *Il diritto nell'era digitale*, a cura di R. Giordano, A. Panzarola, A. Police, S. Preziosi e M. Proto, Milano 2022, 686.

colpa di ogni evento che sia causato (contravvenendo o andando al di là dalla programmazione iniziale) da una IA intrinsecamente "imprevedibile" violerebbe la *ratio* del predetto principio e il già richiamato divieto di responsabilità oggettiva.

In altre parole, il danno da dispositivo intelligente ripropone, accentuandoli, alcuni dei profili più problematici già sorti in relazione alla responsabilità penale per danno da prodotto: l'indebita sovrapposizione tra struttura commissiva e omissiva del reato; l'ostica identificazione dei soggetti personalmente responsabili all'interno delle organizzazioni complesse; l'individuazione del nesso di causalità in relazione a prodotti caratterizzati da opacità; l'accertamento della colpa in situazioni di incertezza scientifica[26].

Nei sistemi di *common law*, invece, la situazione appare parzialmente diversa, atteso che, anche nel diritto penale, la responsabilità oggettiva assume un ruolo importante, nonostante, di recente, si sia riscontrato un suo progressivo "ridimensionamento"[27]. L'imputazione a titolo di colpa fa, difatti, emergere un *liability gap*, rispetto al quale anche la responsabilità penale merita di essere ridiscussa, specie nella prospettiva di dare valore sostanziale ai principi affermati dall'*AI Act*, in particolare il controllo umano, la sicurezza, la *privacy*, la trasparenza, la non discriminazione e il benessere sociale e ambientale.

2. Il testo dell'*AI Act*, invero, si preoccupa di colmare la stessa lacuna definitoria del concetto di intelligenza artificiale, considerando tale quel «sistema automatizzato progettato per funzionare con livelli di autonomia variabili e che può presentare adattabilità dopo la diffusione e che, per obiettivi espliciti o impliciti, deduce dall'input che riceve come generare output quali previsioni, contenuti, raccomandazioni o decisioni che possono influenzare ambienti fisici o virtuali»[28].

[26] Per una disamina più approfondita circa i profili civili e penali della responsabilità di danno da prodotto difettoso derivante da sistemi di intelligenza artificiale, si rinvia a B. Fragasso, *La responsabilità penale del produttore di sistemi di intelligenza artificiale*, in *www.sistemapenale.it*, 13.6.2023.

[27] Cfr. R.A. Duff, *The Realm of Criminal Law*, Oxford 2018, 1 ss.; W.R. LaFave, *Substantive Criminal Law*, Eagan, Minnesota 2018, 572; M.D. Dubber, *An introduction to the Model Penal Code*[2], Oxford 2015, 65, evidenziando che nel *Model Penal Code* questa forma di responsabilità copre ormai solo ipotesi minori, «*è un'opzione solo per gli illeciti civili* sui generis *previsti dal codice, le "violazioni"*». Sicché, ad oggi, non è da escludere che un'imputazione per responsabilità oggettiva non possa comunque finire per dar luogo a qualche problema pure in questi ultimi sistemi, e non già soltanto nei primi.

[28] Cfr. art. 3 della Risoluzione legislativa del Parlamento europeo del 13.3.2024 sulla proposta di regolamento del Parlamento europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale) e modifica alcuni atti legislativi dell'Unione (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)).

È bene chiarire che tale definizione ricomprende sia i sistemi programmati per svolgere uno specifico compito o funzione – incapaci di svolgere attività diverse da quelle per cui sono stati ideati – sia i *General-purpose AI models* (*GPAI*), ovvero sistemi avanzati di portata generale, i quali possono essere utilizzati per molteplici scopi distinti, tra cui la creazione di testi, di video e immagini.

A partire da tale riassetto definitorio, la sfida che il legislatore europeo intende affrontare è proprio quella di individuare specifiche modalità di utilizzo per i sistemi di *GPAI*, atteso che essi, operando su *input* impliciti volti a replicare capacità simili a quelle umane, appaiono difficilmente prevedibili quanto agli esiti decisionali.

Il Regolamento ha, per l'effetto, l'obiettivo di promuovere lo sviluppo e l'utilizzo dell'IA nel pieno rispetto dei diritti fondamentali dell'individuo, nonché della trasparenza. Vengono così individuati limiti e modalità di utilizzo dei *software*, nonché delineate le responsabilità nell'ipotesi in cui si contravvenga a tali requisiti[29].

Ai sensi dell'art. 2, il Regolamento non si applicherà sia ai settori che non rientrano nell'ambito di applicazione del diritto dell'Unione, sia ai sistemi di IA sviluppati o utilizzati con scopi militari e di difesa, oltre che ai sistemi di IA o modelli di IA, ivi compresi i loro *output*, specificamente sviluppati e messi in servizio al solo scopo di ricerca e sviluppo scientifici. Parimenti, esso non troverà il suo campo applicativo con riferimento a soggetti che sfruttino l'IA in attività non professionali.

Le attività di ricerca, prova o sviluppo relative a sistemi di IA o modelli di IA, prima della loro immissione sul mercato o messa in servizio, sono svolte in conformità del diritto dell'Unione applicabile, in quanto il Regolamento lascia impregiudicate le norme stabilite da altri atti giuridici dell'Unione in materia di protezione dei consumatori e di sicurezza dei prodotti.

Per converso, il campo di applicazione delle disposizioni si estende ai fornitori, importatori, distributori, fabbricanti e rappresentanti autorizzati che operano sul mercato europeo, come pure agli utilizzatori, i quali, anche se situati in un paese extra UE, facciano utilizzo di un prodotto del sistema di IA in Europa.

È esclusa, parimenti, l'applicabilità alle autorità pubbliche di un paese terzo e alle organizzazioni internazionali, nell'ipotesi in cui queste ultime sfruttino l'IA nel quadro della cooperazione o di accordi internazionali per la cooperazione, sempre a condizione che tale paese terzo o organizzazione internazionale fornisca garanzie adeguate sul fronte della protezione dei diritti e delle libertà fondamentali delle persone.

[29] Sul piano sanzionatorio, all'art. 99, si prevede, tra l'altro, che, gli Stati membri stabiliscano sanzioni dettagliate e altre misure di applicazione per la violazione dell'*AI Act* da parte degli operatori. Le sanzioni previste devono essere efficaci, proporzionate e dissuasive, oltre che tenere conto degli interessi delle piccole e medie imprese (PMI), comprese le *startup,* e della loro redditività economica.

A titolo esemplificativo, la non conformità al divieto delle pratiche di IA è soggetta a sanzioni amministrative pecuniarie fino a 35 milioni di euro o, se l'autore del reato è un'impresa, fino al 7 % del fatturato mondiale totale annuo dell'esercizio precedente, se superiore. In ipotesi di violazione di altri obblighi, tra cui quelli alla trasparenza, la sanzione amministrativa pecuniaria applicabile arriva sino ai 15 milioni di euro o, se l'autore del reato è un'impresa, fino al 3 % del fatturato mondiale totale annuo dell'esercizio precedente, se superiore. Nel caso di trasmissione di informazioni inesatte, incomplete o fuorvianti agli organismi notificati o alle autorità nazionali competenti, le sanzioni amministrative sono limitate a 7,5 milioni di euro o, se l'autore del reato è un'impresa, fino all'1 % del fatturato mondiale annuo, sempre se superiore.

Nel decidere se infliggere una sanzione amministrativa pecuniaria e nel determinarne l'importo in ogni singolo caso, si dovrà comunque tenere conto di tutte le circostanze pertinenti alla situazione specifica.

L'approccio di questo rinnovato sistema regolatorio è quindi di tipo precauzionale, ossia *risk-based*, preoccupandosi il legislatore di dividere i sistemi di IA in quattro categorie.

Anzitutto, si prevedono sistemi il cui utilizzo determina un rischio c.d. inaccettabile [30] per l'utente. Tali strumenti, atteso che comportano una minaccia considerevole per la persona, sono pertanto vietati. Tra questi, rientrano le IA che utilizzano tecniche di manipolazione comportamentale, ovvero volte a creare un *social scoring*, cioè utilizzate da o per conto di autorità pubbliche, al fine di operare una valutazione di affidabilità degli individui sulla base dei loro comportamenti sociali o caratteristiche personali; parimenti, sono compresi in tale ambito applicativo i sistemi di identificazione biometrica remota e in tempo reale, cioè in grado di operare un riconoscimento facciale sulla base delle immagini raccolte da internet o da filmati di telecamere a circuito chiuso (fa eccezione, in quest'ultimo caso, l'impiego con finalità di *law enforcement*, per la ricerca da parte delle forze di polizia di autori o vittime di specifici reati o per prevenire minacce terroristiche, previa, in ogni caso, autorizzazione dell'autorità giudiziaria).

La seconda categoria è quella degli strumenti ad alto rischio [31], in cui vengono ricondotte le IA che impattano direttamente con la salute, la sicurezza e i diritti fondamentali delle persone. Si pensi, in particolare, alle IA utilizzate per consentire l'accesso ad un servizio pubblico, a quelle funzionali a calcolare il merito creditizio, nonché a quelle usate in fase di *recruiting* per selezionare i *curricula* dei candidati, ai nuovi dispositivi utilizzati in ambito sanitario, o, infine, ai sistemi di guida autonoma utilizzati in veicoli di varia natura [32]. Il commercio e l'utilizzo di questi sistemi è

[30] Cfr. art. 5 della Risoluzione legislativa del Parlamento europeo del 13 marzo 2024 sulla proposta di regolamento del Parlamento europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale) e modifica alcuni atti legislativi dell'Unione (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)).

[31] Cfr. artt. 6 ss. della Risoluzione legislativa del Parlamento europeo del 13 marzo 2024 sulla proposta di regolamento del Parlamento europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale) e modifica alcuni atti legislativi dell'Unione (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)).

[32] Per un'analisi più completa, vedasi, in particolare, l'Allegato III al Regolamento. Merita poi precisare che, in deroga al paragrafo 2, un sistema di IA non è considerato ad alto rischio se non presenta un rischio significativo di danno per la salute, la sicurezza o i diritti fondamentali delle persone fisiche, anche nel senso di non influenzare materialmente il risultato del processo decisionale. È questo il caso quando sono soddisfatte una o più delle seguenti condizioni: a) il sistema di IA è destinato a eseguire un compito procedurale limitato; b) il sistema di IA è destinato a migliorare il risultato di un'attività umana precedentemente completata; c) il sistema di IA è destinato a rilevare schemi decisionali o deviazioni da schemi decisionali precedenti e non è inteso a sostituire o influenzare la valutazione umana precedentemente completata senza un'adeguata revisione umana;

consentito a patto che sia garantito, mediante apposita verifica di conformità *ex ante*, il rispetto di specifici obblighi e requisiti in termini di tutela dei diritti fondamentali, come pure delle regole di trasparenza, di analisi e di gestione dei rischi concreti, che si potrebbero riscontrare durante il loro utilizzo e di *data governance*. È richiesto, altresì, che tali sistemi vengano registrati in un'apposita banca dati UE, ove sia conservata tutta la documentazione tecnica utile a dimostrare la conformità dell'IA a quanto previsto dal regolamento.

La terza categoria concerne i sistemi a rischio limitato[33]. Tra questi rientrano i sistemi di IA – come le *chatbots* – per i quali – atteso il mancato riscontro di rischi specifici – sono previsti semplici obblighi informativi e di trasparenza verso l'utilizzatore finale, tali da consentirgli di apprendere che il contenuto è stato generato dall'IA.

Infine, il regolamento tratta dei sistemi a rischio minimo[34]. Per tali strumenti, l'*AI Act* non prevede alcun obbligo, fatta salva l'autonoma scelta delle imprese di aderire a codici di condotta. Tali codici possono contenere impegni volontari inerenti, a titolo esemplificativo, al divieto di discriminazione di genere, alla sostenibilità ambientale o all'accessibilità da parte delle persone con disabilità.

Per garantire un'attuazione armonizzata del Regolamento, oltre all'istituzione di un apposito Comitato europeo per l'IA, spetterà ai singoli Stati membri nominare Autorità di controllo nazionali con il preciso compito di vigilare sull'effettiva applicazione, da parte degli operatori economici, delle prescrizioni e degli obblighi stabiliti dall'*AI Act*[35].

---

d) il sistema di IA è destinato a eseguire un compito preparatorio per una valutazione pertinente ai fini dei casi d'uso elencati nell'allegato III. Fatto salvo il primo comma, un sistema di IA di cui all'allegato III è sempre considerato ad alto rischio qualora esso effettui profilazione di persone fisiche.

[33] Cfr. art. 52 della Risoluzione legislativa del Parlamento europeo del 13 marzo 2024 sulla proposta di regolamento del Parlamento europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale) e modifica alcuni atti legislativi dell'Unione (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)).

[34] Cfr. art. 69 della Risoluzione legislativa del Parlamento europeo del 13 marzo 2024 sulla proposta di regolamento del Parlamento europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale) e modifica alcuni atti legislativi dell'Unione (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)).

[35] Nel contesto italiano, in una recente segnalazione del Presidente del Garante per la protezione dei dati personali al Parlamento e al Governo, datata 25.03.2024, si segnala la necessità di affidare tale controllo ad un'autorità terza e indipendente. In particolare, il Presidente dell'autorità Garante evidenzia che l'*AI Act* si fonda sull'articolo 16 del TFUE, che è la base giuridica della normativa di protezione dei dati, e che lo stesso Regolamento sull'intelligenza artificiale prevede il controllo delle Autorità di protezione dei dati personali su processi algoritmici che utilizzino dati personali. Si suggerisce, in tal senso, una sinergia tra le due discipline e la loro applicazione da parte di un'unica Autorità, nella prospettiva di fornire effettività alla tutela dei diritti e delle garanzie sanciti nell'*AI Act*.

Inoltre, verrà costituita un'apposita banca dati europea nella quale confluiranno, assieme ai dati rilasciati dai fornitori dell'IA in fase di immissione nel mercato, anche le indicazioni fornite da un'apposita commissione indipendente di esperti.

3. Attesa la summenzionata quadripartizione dei sistemi di IA come proposta nell'ambito dell'*AI Act*, appare opportuno soffermarsi a riflettere circa l'incidenza di tale innovativo approccio *risk-based* nel contesto dell'indagine di responsabilità penale.

Difatti, il reato "da intelligenza artificiale" costituisce solo l'ultima tappa di quella tendenziale crisi del delitto d'evento che caratterizza, ormai da decenni, il diritto penale di fronte alla c.d. "società del rischio"[36] e che, come è noto, interessa molteplici ambiti e assume diversi conseguenti significati[37].

Una prima forma di rilevanza della nozione del rischio può dirsi di tipo politico-criminale, nell'ambito di quello che è stato definito come «il terreno proprio del diritto penale del rischio»[38]. Tale riferimento descrive un modello di ordinamento che si

---

Ad ogni modo, va richiamato il fatto che il Regolamento richiede che tali autorità di vigilanza abbiano competenze che comprendano una conoscenza approfondita delle tecnologie dell'intelligenza artificiale, dei dati utilizzati da queste tecnologie, e dei relativi trattamenti attraverso algoritmi, nonché in materia di protezione dei dati personali, della sicurezza informatica e degli standard esistenti. In caso di rispetto di tali requisiti, l'*AI Act* prevede la possibilità di istituire più autorità, conformemente alle esigenze organizzative dello Stato membro.

[36] Vedasi, sul punto, C. Piergallini, *Il paradigma della colpa nell'età del rischio: prove di resistenza del tipo*, in *RIDPP* 2005, 1684; F. Stella, *Giustizia e modernità. La protezione dell'innocente e la tutela delle vittime*, Milano 2003, 224-235; A. Gargani, *La "flessibilizzazione" giurisprudenziale delle categorie classiche del reato di fronte alle esigenze di controllo penale delle nuove fenomenologie di rischio*, in *LP* 2011, 2, 397 ss.; J.M. Silva Sánchez, *L'espansione del diritto penale. Aspetti della politica criminale nelle società postindustriali*, Milano 2004 (ed. spagn. 1999); F. Herzog, *Società del rischio, diritto penale del rischio, regolazione del rischio*, in *Critica e giustificazione del diritto penale nel cambio di secolo. L'analisi critica della Scuola di Francoforte*, a cura di L. Stortoni-L. Foffani, Milano 2004, 357; F. Centonze, *La normalità dei disastri tecnologici: il problema del congedo dal diritto penale*, Milano 2004. Il concetto di "società del rischio", come noto, si deve a U. Beck, *La società del rischio. Verso una seconda modernità*, Roma 2000 (ed. ted. 1986), che ha notevolmente influenzato le citate riflessioni penalistiche, ponendo l'accento su come le tecniche di produzione moderna creino sistematicamente rischi che la società, nel suo complesso, non è in grado di controllare ed assorbire.

[37] Cfr., *ex multis*, V. Militello, *Diritto penale del rischio e rischi del diritto penale tra scienza e società,* in *Europe in Crisis: Crime, Criminal Justice, and the Way Forward. Essays in honour of Nestor Courakis*, a cura di C. Spinellis, N. Theodorakis, E. Billis, G. Papadimitrakopoulos, Atene 2017, 223-238; C. Brusco, *Rischio e pericolo, rischio consentito e principio di precauzione. La c.d. "flessibilizzazione delle categorie del reato*", in *Criminalia* 2012, 383-414; F. Consulich, voce *Rischio consentito*, in *Enciclopedia del diritto. Reato colposo*, diretto da M. Donini, Milano 2021, 1102 ss., C. Piergallini, *Attività produttive e imputazione per colpa, prove tecniche di "diritto penale del rischio*", in *RIDPP* 1997, 1474 ss.; C. Perini, *Il concetto di rischio nel diritto penale moderno*, Milano 2010.

[38] V. Militello, *Diritto penale del rischio e rischi del diritto penale tra scienza e società*, cit., 2. Si tratta del *Risikostrafrecht* quale "teoria critica del moderno sviluppo del diritto penale", come la descrive L. Kuhlen, *Goltdammers Arkiv fuer Strafrecht*, 1994, 347 (con riferimento ai lavori di C. Prittwitz, *Risiko und Strafrecht*.

distacca dall'illecito costruito su un evento di lesione o di pericolo concreto rispetto a beni individuali afferrabili e dominabili dall'agente tramite relazioni di causa-effetto. Questo utilizzo si presta tanto all'approccio del "diritto penale del nemico"[39], quanto a quello del "diritto penale moderno"[40]. In ambedue questi recenti approcci politico-criminali, si evidenzia il ruolo dominante che il rischio assume per fondare la responsabilità penale nelle società contemporanee, non più caratterizzate – neppure in via prevalente – dalla lesione offensiva dei beni penalmente tutelati.

Un secondo ambito di considerazione penalistica ha natura più propriamente dogmatica. Tale contesto, inquadrabile come «rischio nel diritto penale»[41], rappresenta un criterio normativo utile a restringere la causalità naturalistica, atteso che esso identifica uno dei più solidi criteri su cui si sviluppa la rilevanza penale nell'ipotesi più generale di imputazione obiettiva dell'evento. In tal senso, è ipotizzabile la coesistenza con la lesione o il pericolo concreto all'interno della struttura della previsione incriminatrice, operando il rischio una riduzione teleologica della fattispecie penalmente rilevante.

Una terza forma di rilevanza è, da ultimo, quella che riferisce il rischio come

---

Untersuchungen zur Krise vom Strafrecht und Kriminalpolitk in der Risikogesellschaft, Frankfurt 1993). In un tale diritto penale la funzione centrale diventa il controllo del rischio, più che la repressione del danno. Vedasi, in tal senso, B. Bruenhober, *Von der Unrechtsahndung zur Risikosteurung durch Strafrecht und ihre Schranken*, in *Streitbare Strafrechtswissenschaft*, a cura di F.S. Schuenemann, R. Hefendehl ed al., Berlin 2014, 3 ss.

[39] La prima versione della teoria del "diritto penale del nemico" (*Feindstrafrecht*) fu presentata da Jakobs nel corso delle giornate dei Professori di diritto penale, che hanno avuto luogo a Francoforte nel 1985. È stata illustrata nella sua versione definitiva nel 1999, durante le giornate berlinesi dal titolo "*La scienza del diritto penale a fine millennio*". In estrema sintesi, gli individui sarebbero distinti in persone "giuridiche" (*im recht)* e non; i primi, i *cittadini* (*Bürger*), godrebbero di tutte le garanzie dello Stato di diritto; i secondi, i *nemici (Feinde),* non godrebbero di tali garanzie (o almeno, non godrebbero di tutte le garanzie proprie dello Stato di diritto). Da ciò discenderebbero alcune rilevanti conseguenze: il diritto penale contro il nemico si atteggerebbe a strumento di prevenzione avverso i pericoli o, meglio, contro individui pericolosi; contro costoro si dovrebbe procedere anche prima del fatto delittuoso (in specie, anticipando di molto la soglia dell'intervento penale); in casi di necessità, contro il nemico sarebbe accettabile l'uso della coercizione e finanche della tortura. Cfr. G. Jakobs, *Kriminalisierung im Vorfeld einer Rechtsgutverletzung, in Referat auf der Strafrechtslehrertagung, Frankfurt a.M., im Mai 1985),* in *Zeitschrift für die gesamte Strafrechtwissenschaft*, 97, 1985, 753 ss.; G. Jakobs, *Das Selbstverständnis der Strafrechtswissenschaft vor den Herausforderungen der Gegenwart (Kommentar)*, in *Die Deutsche Strafrechtswissenschaft vor der Jahrtausendwende. Rückbesinnung und Ausblick*, a cura di A. Eser – W. Hassemer-B. Burkhardt, Monaco 2000, 51 ss.

[40] Cfr. W. Hassemer, *Libertà e sicurezza alla luce della politica criminale*, in *Sicurezza e diritto penale*, a cura di M. Donini, M. Pavarini, Bologna 2011, 59 ss.; spec. 71 ss.; 78. Diametralmente opposta è l'opinione di W. Naucke, *La robusta tradizione del diritto penale della prevenzione: illustrazione con intento critico, ivi*, 79 ss. a parere del quale fin dai tempi di Hobbes «il diritto penale della sicurezza […] è uno strumento di potere autoritario, pronto alla violenza». Tale carattere non è mai più venuto meno. «Pertanto è indifferente se esso operi a servizio di una monarchia, di una dittatura o di una democrazia».

[41] V. Militello*, Diritto penale del rischio e rischi del diritto penale tra scienza e società*, cit., 2-3.

componente connaturata allo stesso diritto penale e alla sua unità costitutiva essenziale, il reato, secondo l'impostazione del «rischio del diritto penale»[42]. Ne deriva un'applicazione non tanto al contenuto della *matière penale*, ma all'attività umana posta in essere dall'agente. Il rischio, in tale logica, è quindi riferito all'autore del fatto di reato e non tanto alle caratteristiche dell'offesa realizzata alla vittima.

Sulle distinzioni tra le nozioni di "rischio" e "pericolo", in dottrina, non si è peraltro ancora giunti ad una risposta condivisa[43]. Almeno nella sua formulazione novecentesca[44], l'espressione di rischio non possedeva invero una portata semantica differente rispetto a quella di pericolo, quale giudizio di idoneità causale e di

[42] V. Militello, *Diritto penale del rischio e rischi del diritto penale tra scienza e società*, cit., 3.

[43] Su tutte queste diverse modalità di interazione fra il diritto penale e il rischio si staglia una nozione logica di tale elemento, che è preliminare ad una corretta impostazione dei problemi connessi: per rischio si deve intendere una probabilità di un evento.
Da qui, un problema di chiarimento dei rapporti con la nozione di pericolo, che esprime pure la probabilità di un evento futuro. L'analogia logica non è però identità contenutistica: già a livello generale il rischio si estende a comprendere ogni tipo di evento, tanto negativo (il verificarsi di un danno: ad es. rischio di un sinistro stradale), ma anche – seppur meno di frequente – positivo (ad es. in una lotteria: rischio di vincita).
In ambito penalistico, poi un'analisi attenta a cogliere le diversità sul piano dogmatico fra le due nozioni rileva che il rischio si riferisce ad una valutazione *ex ante*, ma rispetto ad una situazione che è sfociata nella effettiva lesione o danno, e dunque il relativo riferimento serve a rapportare reciprocamente i due stati per verificare se l'evento effettivamente verificatosi rappresenti o meno la concretizzazione del genere di eventi la cui probabilità era stata creata dalla condotta del soggetto. Il pericolo invece esprime una offesa che non si è concretizzata in un risultato lesivo distinto dalla stessa condotta (esempio, il delitto tentato) e dunque rimane una probabilità non confermata dal decorso reale.

[44] Vero è peraltro che in ambito penalistico la nozione di pericolo vanta origini più risalenti ed è penetrata più diffusamente nel lessico dommatico: nel ricostruirne quello che è stato segnalato come il suo "percorso di successo" nel diritto penale tedesco, si è di recente fatto riferimento al suo emergere sin dal XVI secolo (in specie nella *Constitutio Criminalis Carolina* del 1532) in relazione alla legittima difesa (in caso di pericolo inevitabile alla propria integrità). Già all'inizio del XIX secolo il riferimento al pericolo serviva per la sistematica della parte speciale: i manuali cominciavano ad individuare la categoria del pericolo per indicare raggruppamenti nel catalogo dei reati. Ad es., essa compare espressamente nell'intitolazione "dei delitti di comune pericolo" di K. Grolman, e nella contrapposizione fra varie forme di pericolo individuale e pericolo comune nella categorizzazione proposta nel *Lehrbuch* di P.A. Feuerbach. Quest'ultimo autore, peraltro, nel progettare il codice penale Bavarese del 1813, faceva più volte riferimento alla categoria del pericolo, tanto in relazione ai delitti di danno della proprietà connessi ad un pericolo, quanto in relazione alle azioni pericolose che cagionano un danno (considerate come delitti colposi), quanto alla minaccia di un pericolo per la vita come scusante. E la riflessione teorica si dedica al delitto di pericolo già nel 1825, benché la prima trattazione monografica arriverà nel 1886. In proposito, anche per i riferimenti ulteriori, cfr. F.C. Schroeder, *Der Siegeszug der Gefahr im Strafrecht*, in *Gesamte Strafrechtswissenschaft in internationaler Dimension*, a cura di M. Zoeller ed al., Berlin 2013, 247 ss.
Il riferimento al concetto di rischio è invece successivo all'emersione del pericolo, affiancandosi ad esso senza sostituirlo: K. Binding, sin dalla prima edizione di *Die Normen und ihre Uebertretung* del 1876, offre una sistematica del delitto di pericolo ed al contempo introduce la nozione di "rischio moderato, misurato" nel delitto colposo. Da allora il ruolo del rischio nel diritto penale si è andato dilatando, sino ad assumere almeno il triplice significato prima evidenziato. Ma la relazione originaria con il pericolo rimane diffusa e non è raro che i due concetti siano impiegati come sinonimi, il che certo non aiuta a individuare gli specifici ambiti e problemi di volta in volta in considerazione.

propensione al danno fondata su basi scientifiche[45].

La concettualizzazione autonoma si propose solo alla fine degli anni Ottanta, descrivendo una relazione potenziale tra un antecedente e un accadimento, sulla base di una legge scientifica di copertura. Diversamente dal pericolo, tuttavia, esso sarebbe stato riferibile esclusivamente alla condotta e non anche ad un suo effetto[46].

L'elemento differenziale è stato da alcuni individuato sul fronte qualitativo[47]; altra parte della dottrina, al contrario, sostenne la necessità di ripiegare su una caratterizzazione quantitativa, atteso che «il 'pericolo' null'altro è se non un 'rischio' caratterizzato da un'alta possibilità di verificazione del danno all'interesse considerato»[48].

Secondo questa seconda impostazione, conseguentemente, il pericolo si risolverebbe nella "probabilità" o nella "rilevante possibilità" del verificarsi dell'evento dannoso; il rischio resterebbe invece relegato nell'area del mero "possibile", preveduto nel caso di dolo e prevedibile nel caso di colpa[49]. Ne deriva che il rischio si sostanzierebbe in un'ipotesi "diminuita" di pericolo, con una probabilità di

---

[45] Si vedano qui i riferimenti essenziali a C. Piergallini, *Danno da prodotto e responsabilità penale. Profili dommatici e politico-criminali*, cit., 257 ss.; F. Angioni, *Il pericolo concreto come elemento della fattispecie penale. La struttura oggettiva*, Milano 1994, 22 ss., che parla, a proposito della relazione di pericolo, di «causabilità» (*ivi*, 24); G. Fiandaca, *La tipizzazione del pericolo*, in *Dei delitti e delle pene*, 1984, 441; S. Canestrari, *Reato di pericolo*, in *EG*, XXVI, 1991, 1 ss.; M. Gallo, *I reati di pericolo*, in *Foro pen.*, 1969, 1 ss.; A. Gargani, *Il danno qualificato dal pericolo*, Torino 2005, 248; M. Parodi Giusino, *I reati di pericolo tra dogmatica e politica criminale*, Milano 1990, 202 ss., secondo il quale il pericolo è una relazione di probabilità di un risultato futuro da accertare tenendo in considerazione tutti i dati del singolo fatto concreto, mentre la nozione di rischio (non) consentito sarebbe diversa sia per modello di accertamento che per collocazione sistematica. Esso, infatti, fungerebbe da criterio di determinazione della misura oggettiva della colpa e come strumento di imputazione dell'evento nell'ambito della *objektive Zurechnung*. Dal punto di vista dell'accertamento, poi, il superamento del rischio consentito richiederebbe solo una valutazione *ex ante* rapportata alle qualità generiche della condotta, variando peraltro il livello del consenso da parte dell'ordinamento in funzione dell'elemento soggettivo dell'agente. Per una insostituibile sostanza scientifica del pericolo, anche quando astratto, F. D'Alessandro, *Pericolo astratto e limiti soglia. Le promesse non mantenute del diritto penale*, Milano 2012, 172 ss.

[46] V. Militello, *Rischio e responsabilità penale,* Milano 1988, 24 ss., il quale prende le mosse da un'identità strutturale tra le relazioni di pericolo e rischio (*ivi*, 33) e rileva (*ivi*, 21) come il pericolo sia l'unico dei due a figurare nella fattispecie penale come attributo della condotta o di altri elementi (come l'evento).

[47] In questo senso vedasi V. Militello, *Rischio e responsabilità penale, cit*., 17 ss., il quale sostiene che il rischio afferisce alla condotta, mentre il pericolo all'evento.

[48] È questa la definizione di G. Marini, *"Rischio consentito" e tipicità della condotta. Riflessioni*, in *Scritti in memoria di Renato Dell'Andro*, vol. II, Bari 1994, 539 ss. (v. in particolare 542 ss.). Per la soluzione quantitativa propendeva anche C. Perini, *Prospettive del concetto di rischio nel diritto penale moderno*, Garbagnate Milanese 2002. L'Autrice, nel più recente saggio. *Il concetto di rischio nel diritto penale moderno*, Milano 2010, sembra oggi orientata a ritenere i concetti di rischio e pericolo sinonimi (v. 42, 63, 371 ss.)

[49] C. Brusco, *Rischio e pericolo, rischio consentito e principio di precauzione. La c.d. "flessibilizzazione delle categorie del reato"*, cit., 1.

verificazione inferiore[50]. Tale aspetto ha reso forse il rischio oggetto di una tendenziale liceizzazione da parte del legislatore, diversamente dal pericolo, il quale dispone di cautele penalmente sanzionate.

Nel contesto attuale, nondimeno, il rischio ha assunto una connotazione ambigua, assumendo la conformazione di pericolo puramente sospettato, in cui la correlazione tra la condotta e l'evento è incerta e non trova sostegno nella *littera legis*[51]. La contrapposizione con il pericolo come giudizio scientificamente fondato non potrebbe dunque essere più marcata, tanto da segnare l'intero approccio del diritto penale, il quale – secondo questa logica – non potrebbe che riferirsi al rischio se non tramite l'abbandono di ogni riferimento al fatto[52].

In particolare, l'approccio precauzionale determina che, in situazioni di rischio non scientificamente fondate, vi sia uno scostamento dal principio *in dubio pro actione* a quello *in dubio pro omissione*[53].

Attuando il principio di precauzione, di riflesso, il "rischio consentito" appare quale un ossimoro, considerato che ogni rischio in quanto tale deve essere neutralizzato, almeno nei settori socialmente "sensibili", quali quelli della salute pubblica e dell'ambiente[54].

Il timore verso il rischio ha dato così origine, anche nel settore penale, ad un progressivo "azzeramento normativo" dell'incertezza[55]. Ne è emblema, ad esempio, la colpa di organizzazione, in quanto intesa come rimprovero collettivo di una scorretta gestione del rischio-reato.

---

[50] G. Marini, *«Rischio consentito» e tipicità della condotta. Riflessioni*, cit., 544.

[51] Sul rischio come giudizio di mero sospetto non scientificamente fondato cfr. C. Piergallini, *Danno da prodotto e responsabilità penale. Profili dommatici e politico-criminali*, cit., 520.

[52] Sulla caratterizzazione causale del rischio, invece, vedasi G. Marini, *«Rischio consentito» e tipicità della condotta. Riflessioni*, cit., 565.

[53] Per un inquadramento del principio di precauzione e le ricadute penalistiche derivatene, si rinvia a F. Consulich, *Tutela del consumatore*, in *Commentario breve alle leggi penali complementari*, a cura di F. Palazzo-C.E. Paliero, Padova 2007, 2977 ss. Sul modello di diritto punitivo incentrato sulla precauzione come paradigma alternativo all'impiego del pericolo, anche quando astratto, cfr. M. Donini, *Diritto penale «classico» e diritto penale «moderno»*, in Id., *Il volto attuale dell'illecito penale*, Milano 2004, 120. Si vedano, sul rapporto tra principio di precauzione e diritto penale, gli studi monografici di D. Castronuovo, *Principio di precauzione e diritto penale*, Roma 2012, 43 ss.; F. Consorte, *Tutela penale e principio di precauzione*, Torino 2013, 63 ss., 79 ss., 180 ss.; E. Corn, *Il principio di precauzione nel diritto penale. Studio sui limiti all'anticipazione della tutela penale*, Torino 2013, 27 ss.

[54] Sul rischio come modello del diritto penale della postmodernità, vedasi G. De Francesco, *Pericolo, rischio, incertezza. Il controllo penale e i suoi confini nella temperie della postmodernità*, in *Regole dell'agricoltura. Regole del cibo. Produzione agricola, sicurezza alimentare e tutela del consumatore*, a cura di E. Sirsi e M. Goldoni, Pisa 2005, 125 ss.

[55] F. Consulich, voce *Rischio consentito*, cit., 1107.

Non deve quindi stupire che – nell'ambito di quello che Federico Stella definiva "*shock* da modernità"[56] – si sia tentato – pur non senza critiche[57] – di elidere il rischio conseguente all'utilizzo dei sistemi di IA, riconducendo la responsabilità penale "diretta" della macchina nell'alveo dei modelli di *corporate liability*[58].

La traccia segnata dall'*AI Act* di recente adozione, nondimeno, fa riemergere nell'ambito dell'indagine di responsabilità conseguente all'utilizzo di sistemi di IA la rilevanza dell'analisi del rischio proprio del contesto in cui opera la macchina. L'approccio *risk-based,* caratteristico di tale regolamentazione, infatti, sottende che un qualche grado di rischio, con riferimento all'intelligenza artificiale, sia sempre presente e dunque "ineliminabile".

Ne consegue che anche la responsabilità penale, specie in tale peculiare ambito, deve saper tenere conto – nella prospettiva di fornire risposte effettive e adeguate – di quella quota di rischi che non vengono limitati, né tantomeno esclusi, dalle cautele tecnicamente possibili ed economicamente sostenibili, senza privare, a priori, di utilità sociale l'attività pericolosa. Ciò che rimane, secondo un meccanismo di eliminazione "in negativo" dal pericolo di danno di un determinato fatto, è appunto il "rischio consentito"[59], ossia quello che in ogni caso sopravvive all'applicazione dei presidi imposti dal diritto[60], rischio peraltro insito in ogni attività "umana", per sua natura fallace.

Il rischio "da intelligenza artificiale", nondimeno – così come prospettato *dall'AI Act* – è un rischio peculiare, che, come si vedrà meglio in seguito, rappresenta un mero "*legal risk*"[61] anziché un rischio basato sulla realtà del fatto concreto. Tale prospettiva

[56] F. Stella, *Giustizia e modernità. La protezione dell'innocente e la tutela delle vittime*, cit., 292-293.
[57] A. Cappellini, *Machina delinquere non potest? Brevi appunti su intelligenza artificiale e responsabilità penale*, cit., 13 ss.
[58] G. Hallevy, *The Criminal Liability of Artificial Intelligence Entities – from Science Fiction to Legal Social Control*, in *Akron Intellectual Property Journal* 2010, 171 ss. Ciò che impedisce la breccia nel *machina delinquere non potest*, per Hallevy, è dunque un pregiudizio antropocentrico e metafisico.
[59] In realtà, la nozione di rischio consentito ha due possibili letture: quella che lo accosta a tutte le attività pericolose ma socialmente ammesse e quella che lo impiega per definire il rischio residuale che sopravvive all'applicazione di tutte le cautele possibili e dunque ammesso semplicemente perché ineliminabile, naturalmente ove la norma giuridica non lo vieti all'esito di un bilanciamento costi-benefici (qui il criterio dello scopo della norma violata assume una possibile utilità). Sulla duplicità di significati, vedasi, G. Forti, *Colpa ed evento nel diritto penale*, Milano 1990, 456. Sulla nozione di rischio consentito come rischio generale della vita quotidiana o rischio residuale tollerato, si veda W. Frisch, *Zum gegenwärtigen Stand der Diskussion und zur Problematik der obiektiven Zurechnungslehre*, in *G.A. StR.* 2003, 723; nello stesso senso C. Roxin e L. Greco, *Strafrecht, Allgemeiner Teil*[5], I, München 2020, 488.
[60] F. Consulich, voce *Rischio consentito*, cit., 1108.
[61] Per un approfondimento in tema di "*legal risk*", vedasi T. Mahler, *Defining legal risk*, Turku 2007.

– che quasi avvicina il "rischio europeo" al concetto di "pericolo" per la lesione dei diritti più che alla probabilità circa il verificarsi di un evento di danno – comporta, difatti, alcune ulteriori considerazioni.

4. L'individuazione di diversi livelli di rischio, come proposta dall'*AI Act*, impone, così, anche per il penalista, la necessità di ricondurre la riflessione circa l'indagine di responsabilità da intelligenza artificiale nell'alveo di un sistema di *risk assessment*, atteso che – come è evidente – l'*AI Act* "normativizza" una serie di disposizioni di natura cautelare[62].

Infatti, il "rischio consentito" non è rilevabile guardando al tipo di pericolo in sé, ma dipende dalla qualificazione dell'attività da cui esso scaturisce, in rapporto alle cautele che l'ordinamento ha imposto ad essa alla luce di considerazioni di utilità generale.

Invero, la teoria del c.d. rischio consentito – cui, in tale contesto, si intende fare riferimento, senza pretese di esaustività – è stata elaborata da Roxin[63] con precipuo riferimento ai reati colposi, ma fu poi estesa a quelli dolosi, interessandosi di definire un livello oggettivo di rischio lecito generale, il cui superamento determina conseguenze imputative.

Tale concetto richiama quello di adeguatezza sociale, tematizzato in precedenza come criterio di discernimento tra rischi accettabili e vietati. L'azione socialmente adeguata, secondo tale ricostruzione, si caratterizza quale prototipo di azione produttiva di un rischio consentito[64]. Nell'ambito della teoria dell'adeguatezza sociale si prendeva così atto della necessità di autorizzare una molteplicità di rischi per consentire attività indispensabili alla vita associata, sicché compito della fattispecie penale era selezionarne solo alcuni, identificando il disvalore dell'azione commessa[65].

Il rischio consentito rappresenta così un "anfibio giuridico"[66] di collegamento tra il fatto e il diritto, ossia tra il rischio e il consenso, assumendo dapprima una dimensione empirica e, successivamente, valutativa, quale meccanismo di ausilio alla

[62] R. Compostella, *Auto a guida autonoma e diritto penale. Profili di responsabilità individuale e collettiva*, cit., 161.
[63] Si veda C. Roxin e L. Greco, *Strafrecht*, *Allgemeiner Teil*, cit., 488, secondo cui «*Da die einhaltung des erlaubten Risikos eine Zurechnung zum objektiven Tatbestand verhindert, ist also die Verursachung einer Rechtsgüterverletzung, die trotz Beachtung aller Verkehrsregeln zustande kommt, keine Tatbestandshandlung. Das gilt für fahrlässige und vorsätzliche Delikte gleichermaßen*».
[64] Lo rileva già S. Fiore, *Cause di giustificazione e fatti colposi*, Padova 1996, 141 ss.
[65] C. Fiore, *L'azione socialmente adeguata*, Napoli 1966, 187 ss.
[66] F. Consulich, voce *Rischio consentito*, cit., 1105.

qualificazione tipica o atipica di una fattispecie concreta[67].

Infatti, nonostante alcuni facciano operare l'istituto sul piano dell'elisione di antigiuridicità[68] ovvero sull'integrazione di una scriminante atipica[69], la dottrina prevalente riconduce la riflessione sul rischio consentito in un campo autonomo della tipicità[70], diverso dalla causalità, accomunandola, sotto questo profilo, tanto alla teoria dell'imputazione oggettiva, quanto alla concezione normativa della colpa[71].

Di fatto, ad oggi – nella *matière penale* – il rischio consentito è relegato a statuto meramente argomentativo, nel contesto di una concezione non legalistica della colpa, che al crescere delle pretese preventive – soprattutto nelle organizzazioni e nelle attività complesse – non fa corrispondere un analogo tasso di prevedibilità della prudenza dovuta, spesso ricavata *ex post* rispetto alla verificazione dell'evento avverso.

L'utilità del concetto è dunque tuttora limitata al piano dell'interpretazione della fattispecie e, dal punto di vista politico-criminale, ai profili di legittimazione delle attività pericolose e alla discrasia che esse presentano tra offesa sostanziale e liceità formale della condotta dell'agente[72].

L'*AI Act*, come già rilevato, – pur non preoccupandosi di definire i possibili diversi profili di responsabilità penale riconnessi alla materia – presuppone un approccio *risk-based*. Se, difatti, le cautele e gli *standard* richiesti per l'immissione nel mercato e l'utilizzo di strumentazioni che si avvalgono di IA sono consentiti in ragione del rischio

[67] Sulla relazione tra rischio consentito e adeguatezza sociale vedasi A. Perin, *Prudenza, dovere di conoscenza e norma penale. Proposta per un metodo di giudizio*, Napoli 2020, 213. Il primo e principale teorizzatore dell'adeguatezza sociale è sicuramente H. Welzel, *Studien zum System des Strafrechts*, in *ZStW*, 58, 1939, 491 ss.; Id., *Das deutsche Strafrecht*, Berlino 1967, 56, che ne prospettava l'applicazione al normale traffico negoziale. Non a caso con riferimento al carattere permesso o vietato del rischio opera ancora un'associazione con il carattere socialmente adeguato del comportamento U. Murmann, *Zum Tatbestand der Beihilfe*, in *Jus* 1999, 52. La teoria ha trovato risonanza in Italia grazie a C. Fiore, *L'azione socialmente adeguata*, cit., 169 ss. e, sulla collocazione della adeguatezza sociale nella sistematica del reato, 186 ss.

[68] Si veda la ricostruzione di V. Militello, *Rischio e responsabilità penale*, cit., 61. Più di recente, ribadisce l'inquadramento, nell'ottica di assicurare un controllo di proporzionalità in ordine agli interessi in conflitto, A. Cavaliere, *L'errore sulle scriminanti nella teoria dell'illecito penale. Contributo ad una sistematica teleologica*, Napoli 2000, 437 ss.

[69] S. Fiore, *Cause di giustificazione e fatti colposi*, cit., 58 ss.

[70] La dottrina tedesca tradizionalmente inquadra il rischio consentito come parametro di esclusione della tipicità della condotta nell'ambito dei reati di evento: cfr. D. Kienapfel, *Das Erlaubte Risiko im Strafrecht. Zur Lehre vom sozialen Handlungsbegriff*, Frankfurt am Main 1966, 14. Agire nell'ambito del rischio consentito attribuisce all'agente un beneficio derivante dalla tolleranza dell'ordinamento per il tipo di eventi dannosi che ne dovessero pur causalmente conseguire. Cfr. I. Puppe, *Kausalität der Sorgfaltspflichtverletzung*, in *Jus,* 1982, 661, che parla di *Benefiz des erlaubten Risikos*. Quanto alla recente dottrina italiana in materia, vedasi F. Consulich, voce *Rischio consentito*, cit., 1111.

[71] C. Piergallini, *Colpa (diritto penale)*, in *ED*, *Annali* X, Milano 2017, 250 ss.

[72] F. Consulich, voce *Rischio consentito*, cit., 1135.

proprio di questi sistemi, ne deriva che anche l'analisi della responsabilità penale dovrà tenere conto di quello che è – normativamente – un rischio ineliminabile, predefinito e "socialmente adeguato".

In definitiva, nel caso delle regole cautelari dotate di portata preventiva non totale – come quelle predisposte nel contesto dell'intelligenza artificiale, ove si sottende l'ineliminabilità di forme di rischio lesive di diritti fondamentali – quando l'evento si realizzi nonostante l'osservanza della diligenza richiesta, si realizzerà una necessaria allocazione collettiva del rischio. Il *Restrisiko*, inevitabile per la conservazione dell'utilità derivante dall'attività rischiosa, viene "assunto" dall'ordinamento, che procederà eventualmente con misure perequative a danno delle vittime dell'offesa[73].

D'altra parte, questo approccio basato sul rischio consentito postula un interesse meritevole di tutela alla cui implementazione è funzionale la condotta pericolosa; ciò evidentemente non accade quando il soggetto strumentalizza la sua azione ad un'offesa, non compensata da alcun correlato vantaggio.

Ne consegue che, nelle aree in cui tali tecnologie altamente autonome sono ritenute accettabili e i rischi tutto sommato tollerabili – in quanto controbilanciati dai grandi benefici ricavabili dal loro utilizzo –, all'interno cioè dell'area di "rischio socialmente consentito" delineata da ciascun sistema[74], si dovrebbe verosimilmente riconoscere che di norma non dovrebbe sorgere alcun tipo di responsabilità penale per i programmatori, i produttori o gli utenti. Conseguentemente, ove si eccedano i confini della prevedibilità del danno pur nel rispetto degli *standards* precauzionali richiesti[75], gli eventuali e rari eventi sfortunati occorsi dovrebbero verosimilmente restare vieppiù

[73] Ciò accade a meno che naturalmente non si verifichi il frequente meccanismo di subduzione della colpa specifica in una colpa generica, cioè di fronte all'inefficienza della cautela specifica si palesi una più generica norma di cautela, perennemente in concorrenza con la prima e idonea allo scopo preventivo. Sulla possibilità che la colpa generica subentri a quella specifica in ragione dell'obsolescenza della prima o dell'inadeguatezza alle circostanze, cfr. G. Marinucci, *La colpa per inosservanza di leggi*, Milano 1965, 198. Sulla interscambiabilità giurisprudenziale tra colpa specifica, generica, monosoggettiva, plurisoggettiva, vedasi F. Donelli, *La violenza come infortunio: specificare la colpa e ragionare per principi*, in *www.lalegislazionepenale.eu*, 24.4.2021, 7.

[74] Ossia, quel rischio residuale al rispetto degli *standards* imposti dall'*AI Act*. Sarebbe ancora meglio se quest'area di rischio venisse delineata a livello globale nel modo più uniforme possibile, al fine di evitare le discrepanze nella classificazione dei casi *cross-border* che potrebbero porsi, come osservato da V. Manes, *L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia*, in *Intelligenza artificiale. Il diritto, i diritti, l'etica*, a cura di U. Ruffolo, Milano 2020, 548 ss.

[75] In tale ipotesi, sarebbe possibile rinvenire uno spazio per l'attuazione del modello della *corporate criminal liability*, applicando sanzioni penali all'ente o per la colpa di organizzazione ovvero, indipendentemente, come mera conseguenza della mancata adozione delle cautele richieste. Si veda, più nel dettaglio, V. Mongillo, *Corporate criminal liability for AI-related crimes: possible techniques and obstacles*, in *Traditional Criminal Law Categories and AI: Crisis or Palingenesis?*, a cura di L. Picotti e B. Panattoni, cit., 82 ss.

appannaggio dei meccanismi propri della responsabilità civile [76], secondo l'impostazione propria delle attività pericolose lecite[77].

Queste finalità costituiscono il presupposto della classificazione in categorie di "rischio consentito", contenute nell'*AI Act*, che si propone, quindi, di definire a monte quel necessario equo bilanciamento tra benefici e rischi, in grado di assicurare, in altri termini, lo sviluppo, l'uso e l'adozione dell'intelligenza artificiale nel mercato interno, senza pregiudicare la protezione, la tutela e la salvaguardia dei diritti fondamentali dell'Unione.

5. Questo approccio alla tematica assunto dall'*AI Act*, rappresentando la prima regolamentazione europea di carattere generale in tema di intelligenza artificiale, mostra ancora una volta come la tutela dei diritti fondamentali si stia sviluppando, sempre di più, in una logica multilivello e "orizzontale"[78], che vede la persona, prima dei "beni giuridici", al centro della tutela[79].

La sovrapposizione di differenti "*layers*" nell'ambito del riconoscimento dei diritti umani – denotativi di un assetto costituzionale "culturalmente europeo" – impone, di riflesso, l'attuazione della massima espansione possibile delle garanzie.

Specificatamente, l'uniformità degli *standards* cautelari richiesti nell'ambito della commercializzazione e dell'utilizzo degli strumenti di IA – oggetto della recente regolamentazione – evidenzia come l'effettività del rispetto dei diritti umani debba essere individuato, in modo omogeneo, anche nell'ambito delle tutele giuridiche

[76] Vedasi, in questo senso, M.E. Florio, *Il dibattito sulla responsabilità penale diretta delle IA: "molto rumore per nulla"?*, cit., 17.

[77] La penale responsabilità, in tale ambito, non può infatti prescindere dall'accertamento di una posizione di garanzia e dalla necessaria connessione della violazione cautelare con l'evento lesivo. Più in particolare, si rimanda a D. Notaro, *Le insidie della colpa nella gestione di attività pericolose lecite. La predisposizione delle pratiche ludico-sportive*, in *Criminalia* 2020, 8 ss.

[78] Con l'espressione "tutela multilivello dei diritti" sul fronte europeo e internazionale s'intende designare, dunque, «*il complesso di istituti, tanto di origine normativa che giurisprudenziale, attraverso cui si articolano le competenze e le relazioni tra le varie istanze giurisdizionali degli ordinamenti nazionali e sovranazionali davanti a cui è possibile far valere la tutela dei diritti fondamentali*». Cfr., in materia, A. Cardone, voce *Diritti fondamentali (tutela multilivello)*, in *ED, Annali* IV, Milano 2011, 336 ss.

[79] Per un approfondimento sul tema, si rimanda a F. Viganò, *Diritto penale e diritti della persona*, in *www.sistemapenale.it*, 13.3.2023, 1 ss. Il contributo dell'autorevole giurista è pubblicato anche nel volume collettaneo *Studi in onore di Carlo Enrico Paliero*, a cura di G. Mannozzi, C. Perini, M.M. Scoletta, C. Sotis e S.B. Taverriti, Milano 2022, 845 ss. In chiusura del saggio, l'eminente penalista evidenzia, in modo critico, che «*l'attenzione dello studioso italiano è sempre più attratta dai reati a vittima diffusa, che offendono beni collettivi e che coinvolgono grandi organizzazioni complesse; e finisce così per trascurare la realtà quotidiana della prassi penalistica, che continua a essere popolata di vicende criminose scaturenti da conflitti interpersonali*» tra singoli.

disponibili.

L'*AI Act* pone difatti l'interprete di fronte al tessuto di uno *ius commune* che – piuttosto che caratterizzarsi per la sovrapposizione o il confronto tra discipline positive – appare formato da una trama di principi generali di ordine sovranazionale. Tale nuova "Costituzione dell'UE sull'IA" [80] dovrà necessariamente intrecciarsi all'ordito dei diritti nazionali, secondo una «logica di integrazione ed assimilazione, che si riversa in particolare nella tematica dei diritti fondamentali»[81].

I principi propri della *matière penale*, pur senza perdere la loro caratterizzazione, si arricchiscono così – nel contesto dell'IA – del bilanciamento con "nuovi valori", quali quelli della *privacy*, della trasparenza delle informazioni, del controllo umano, della sicurezza e della non discriminazione[82].

In definitiva, stante l'approccio di tipo *risk-based* in cui si innestano tali assunti, i parametri di indagine della responsabilità dovranno sapersi "riadattare" all'intento – proprio dell'*AI Act* – di "regolare" e normare rischi per loro natura ineliminabili.

Le categorie istituzionali penalistiche – in un settore dove l'attuazione rigorosa del principio di legalità e di colpevolezza denota la presenza di un *liability gap* – dovranno così "cedere il passo" a differenti forme di tutela, idonee comunque ad assicurare sanzioni "proporzionate ed effettive" alla lesione dei diritti fondamentali, in linea con le considerazioni elaborate dalla giurisprudenza sovranazionale[83].

6. La categorizzazione fatta propria dall'*AI Act* – qui descritta necessariamente in modo sintetico – sconta peraltro i limiti propri di un approccio "statico" al fenomeno dell'intelligenza artificiale, che appare limitante nell'affrontare le continue sfide imposte dalla rapida evoluzione del progresso tecnologico. In altri termini, si pone la necessità evidente di attuare un approccio maggiormente flessibile, proporzionale e

[80] C. Novelli, *L'Artificial Intelligence Act Europeo: alcune questioni di implementazione*, cit., 95.
[81] F. Patroni Griffi, *Corti nazionali e Corti europee: un problema di confini?*, in *federalismi.it*, 3, 2021, 30 ss. Cfr. anche M.R. Ferrarese, *I confini e la voglia di attraversarli,* in *Attraversare i confini del diritto*, a cura di L. Torchia, Bologna 2016, 55 ss.
[82] S. Quattrocolo, *Intelligenza artificiale e giustizia: nella cornice della Carta Etica Europea, gli spunti per un'urgente discussione tra scienze penali e informatiche*, cit., 10 ss.
[83] La giurisprudenza sovranazionale di Strasburgo, infatti, in materia di responsabilità medica, ha escluso la necessaria previsione di sanzioni penali da parte degli Stati in ipotesi di negligenza sanitaria, purché sia assicurata l'effettiva della tutela del diritto alla vita di cui all'art. 8 CEDU. Vedasi, per una disamina in materia, D. Pranka, *The price of medical negligence – Should it be judged by the criminal court in the context of the jurisprudence of the European Court of Human Rights*?, Kaunas 2021, 124 ss.

"semi-quantitativo"[84], funzionale a consentire un aggiornamento e una valutazione "*case by case*" delle stesse categorie di rischio, così come proposte.

Infatti, se da un lato la nuova regolamentazione si preoccupa di inquadrare gli strumenti di IA nel contesto di un bilanciamento a priori tra benefici dell'IA e rischio per la lesione dei diritti fondamentali, non appare possibile rinvenire modi e termini che possano consentire un processo di aggiustamento e di adattamento in grado di considerare la pluralità multifattoriale delle questioni coinvolte.

L'intelligenza artificiale è, cioè, considerata dall'*AI Act* quale mero "prodotto", in linea con la regolamentazione europea in tema di *EU product safety legislation*[85]. Parimenti, l'*AI Act* non tiene conto dell'interazione e interconnessione tra le fonti di pericolo, i profili di vulnerabilità e i valori esposti, ma li tratta come isolati *standard* tecnici. Ciò si accompagna, peraltro, alla mancanza di un giudizio di proporzionalità tra le misure di mitigazione del rischio e i principi e i diritti coinvolti. Conseguentemente, secondo il legislatore europeo, l'impatto che l'IA può avere sui valori e i diritti fondamentali europei appare predeterminato.

Le criticità che emergono dal testo dell'*AI Act*, oltretutto, attengono alla natura del rischio considerato. Infatti, ciò che appare rilevante è un mero "*legal risk*"[86] – che esprime il possibile danno derivante dalla violazione di regole imposte a priori – più che un rischio che si fonda sulla realtà del fatto concreto. In aggiunta, il modello di valutazione adottato dal legislatore europeo non soddisfa neppure la natura distintiva del "*legal risk*", atteso che non valuta in modo comparativo e proporzionato il peso specifico delle norme giuridiche. Al contrario, la valutazione del rischio nell'*AIA* appare impostata come un modello matematico che tratta le norme giuridiche quali *standards*, il cui rispetto o meno è demandato ad una valutazione esclusivamente tecnica[87].

Conseguentemente, il rischio è categorizzato attraverso un rigido elenco di ambiti delle IA potenzialmente dannosi per i principi e i diritti fondamentali. La valutazione,

[84] Per una disamina sul punto, vedasi il recente contributo di C. Novelli, F. Casolari, A. Rotolo, M. Taddeo, L. Floridi, *AI Risk Assessment: A Scenario-Based, Proportional Methodology for the AI Act*, in *Digital Society* 2024, 18 ss.

[85] Tuttavia, esso è compatibile con uno *standard* di gestione del rischio utilizzato in settori ad alta sicurezza in altri ordinamenti. Un esempio ne è il principio adottato dal Regno Unito del "*As Low As Reasonably Practicable*" (*ALARP*).

[86] Per un approfondimento in tema di "*legal risk*", vedasi T. Mahler, *Defining legal risk*, cit.

[87] N. Smuha, E. Ahmed-Rengers, A. Harkens, L. Wenlong, J. Maclaren, R. Piselli, K. Yeung, *How the EU can achieve legally trustworthy AI: A response to the European commission's proposal for an artificial intelligence act*, in *Artificial Intelligence - Law, Policy, & Ethics eJournal,* August 2021.

nondimeno, non può essere uno strumento neutrale: essa riflette l'inclinazione al rischio di una comunità specifica in un determinato contesto storico e culturale, bilanciando i costi e i benefici della mitigazione dello stesso in modo dinamico e diacronico[88]. Diversamente opinando, si rischierebbe di promuoverne un mero "valore normativo", a discapito di altri valori altrettanto fondamentali.

Questa riflessione comporta che le misure di gestione del rischio dovrebbero essere modulate solo all'esito di un effettivo processo di bilanciamento, adattabile e coerente allo stato dei fatti.

Infatti, nonostante esso rivendichi di essere informato al compromesso tra l'interesse allo sviluppo economico e la protezione dei diritti fondamentali, l'*AIA*, in concreto, predetermina il giudizio di proporzionalità che risolve l'interferenza tra valori. Inoltre, non solo l'elenco dei diritti fondamentali protetti dalla proposta è particolarmente ricco, ma include anche diritti interrelati, rendendo così difficile un bilanciamento orizzontale tra diritti fondamentali concorrenti.

Il difetto del modello, peraltro, non riguarda solo la mancanza di granularità nell'analisi dei valori e dei diritti. L'*AIA* manca anche di una rappresentazione accurata delle fonti di pericolo delle IA, di ciò che rende le persone vulnerabili di fronte a questi pericoli e della valutazione circa la preesistenza di meccanismi, anche normativi, di mitigazione dei medesimi.

Per migliorare l'attuazione dell'*AIA*, una parte della più recente dottrina[89] ha quindi proposto una metodologia di valutazione che includa molteplici fattori, unitamente alle loro interferenze, in modo da fornire un giudizio di proporzionalità idoneo a riesaminare le categorie individuate. Tale impostazione, ad ogni modo, andrebbe ad integrarsi e non a sostituirsi agli intervalli di tolleranza prospettati dall'*AIA*. Al contrario, potrebbe essere presa in considerazione la possibilità di applicare le quattro categorie, in una logica "orizzontale", a ciascuna delle IA elencate nell'*AIA*, in modo che, in condizioni variabili – ad esempio, un'interferenza specifica tra i diritti fondamentali coinvolti – lo stesso sistema possa essere trattato come inaccettabile, ad alto rischio, a rischio limitato o a rischio minimo. In tal modo, le categorie non dipenderebbero di *default* dagli ambiti delle IA, ma dagli scenari del mondo reale associati all'applicazione dei sistemi di IA, alla luce dell'incidenza e della combinazione

[88] J.R. Krebs, *Risk, uncertainty and regulation*, in *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 369 (1956), 2011, 4842–4852.
[89] C. Novelli, F. Casolari, A. Rotolo, M. Taddeo, L. Floridi, *AI Risk Assessment: A Scenario-Based, Proportional Methodology for the AI Act*, cit., 9 ss.

multifattoriale delle cause di rischio.

Tale approccio, definibile come "semi-quantitativo", si articola in due fasi: la costruzione degli scenari di rischio e la valutazione quantitativa basata sulla proporzionalità. Per la costruzione degli scenari, il riferimento utilizzato è al quadro teorico dell'*Intergovernmental Panel on Climate Change (*IPCC)[90] e alla letteratura sui rischi legati ai cambiamenti climatici[91]. Per la seconda fase, si è avuto riguardo al metodo quantitativo, come sviluppato da parte della dottrina per bilanciare i principi giuridici[92]. Quest'ultima fase valutativa mira a verificare se la categoria assegnata a seguito della costruzione dello scenario sia o meno proporzionata ai valori coinvolti nell'impiego delle IA.

Specificatamente, questa metodologia "semi-quantitativa" può avere una triplice applicazione nel contesto dell'*AIA*[93].

In primo luogo, supporta l'implementazione orizzontale dell'*AIA* introducendo un modello basato sugli scenari, il quale categorizzerebbe i sistemi di IA in modo flessibile – inaccettabili, ad alto rischio, a rischio limitato o a rischio minimo – a seconda delle situazioni specifiche.

---

[90] Per costruire scenari di rischio, l'*Intergovernmental Panel on Climate Change* (IPCC) fornisce un modello di valutazione del rischio multifacetico, che è poi stato affinato dalla successiva letteratura e che si può utilizzare per valutare i rischi delle IA. Le magnitudini dei rischi associate sia al cambiamento climatico che all'AI sono influenzate da una serie di fattori interagenti, che producono risultati dipendenti dal contesto.
Il quadro riconosciuto per valutare i compromessi insiti nella formulazione di strategie di mitigazione del rischio è stato spesso concepito dall'IPCC per valutare i rischi legati ai cambiamenti climatici, ad esempio, il rischio di catastrofi, come la conseguenza di tre determinanti: pericolo (*hazard*), esposizione (*exposure*) e vulnerabilità (*vulnerability*). In generale, il pericolo si riferisce alle fonti di potenziali effetti avversi sugli elementi esposti; l'esposizione si riferisce all'inventario degli elementi all'interno del raggio d'azione della fonte di pericolo; la vulnerabilità si riferisce all'insieme di attributi o circostanze che rendono gli elementi esposti suscettibili agli effetti avversi quando impattano sulla fonte di pericolo. L'approccio dell'IPCC può essere ulteriormente sviluppato, come nel quadro per la valutazione del rischio dei cambiamenti climatici proposto da Simpson, il quale valuta il rischio a un livello più basso di astrazione includendo i singoli componenti dei determinanti del rischio, cioè i *drive*r. Tale dottrina amplia l'approccio dell'IPCC incorporando un quarto determinante del rischio: la risposta (*response*), che si riferisce alle misure esistenti che contrastano o mitigano il rischio. Inoltre, contestualizzano la valutazione del rischio includendo diversi tipi di rischio con i propri determinanti. Così, secondo tale quadro, il rischio complessivo deriva dall'interazione tra determinanti, *driver* e tipi di rischio. Questi tre insiemi di relazioni si verificano in fasi di crescente complessità. L'*AIA* considera solo la fase di minore complessità, dove i fattori di rischio rilevanti sono i determinanti presi staticamente, trascurando quindi le interazioni tra i loro driver (o con tipi di rischio trasversali).

[91] N.P. Simpson, K.J. Mach, A. Constable, J. Hess, R. Hogarth, M. Howden, J. Lawrence, et al., *A framework for complex climate change risk assessment,* in *One Earth*, 4 (4), 2021, 489–501.

[92] Vedasi R. Alexy, *A theory of constitutional rights*, Oxford 2002; Id., *On balancing and subsumption. A structural comparison*, in *Ratio Juris*, 16 (4), 2003, 433–449.

[93] C. Novelli, F. Casolari, A. Rotolo, M. Taddeo, L. Floridi, *AI Risk Assessment: A Scenario-Based, Proportional Methodology for the AI Act*, cit., 22 ss.

In secondo luogo, agevolerebbe la valutazione della significatività del rischio, fornendo ai soggetti che implementano le IA l'opportunità di riesaminare i livelli di rischio dei loro sistemi, determinando un ridotto onere regolatorio.

Infine, si faciliterebbe e armonizzerebbe l'implementazione del sistema interno di gestione del rischio come delineato nell'articolo 9 dell'*AIA*[94], offrendo un quadro regolatorio più completo, proteso ad aiutare i soggetti che implementano le IA

[94] L'art. 9, rubricato «*Risk management system*», dispone quanto segue:
«*1.A risk management system shall be established, implemented, documented and maintained in relation to high-risk AI systems.*
2.The risk management system shall consist of a continuous iterative process run throughout the entire lifecycle of a high-risk AI system, requiring regular systematic updating. It shall comprise the following steps:
(a) identification and analysis of the known and foreseeable risks associated with each high-risk AI system;
(b) estimation and evaluation of the risks that may emerge when the high-risk AI system is used in accordance with its intended purpose and under conditions of reasonably foreseeable misuse;
(c) evaluation of other possibly arising risks based on the analysis of data gathered from the post-market monitoring system referred to in Article 61;
(d) adoption of suitable risk management measures in accordance with the provisions of the following paragraphs.
3. The risk management measures referred to in paragraph 2, point (d) shall give due consideration to the effects and possible interactions resulting from the combined application of the requirements set out in this Chapter 2. They shall take into account the generally acknowledged state of the art, including as reflected in relevant harmonised standards or common specifications.
4. The risk management measures referred to in paragraph 2, point (d) shall be such that any residual risk associated with each hazard as well as the overall residual risk of the high-risk AI systems is judged acceptable, provided that the high-risk AI system is used in accordance with its intended purpose or under conditions of reasonably foreseeable misuse. Those residual risks shall be communicated to the user.
In identifying the most appropriate risk management measures, the following shall be ensured:
(a) elimination or reduction of risks as far as possible through adequate design and development;
(b) where appropriate, implementation of adequate mitigation and control measures in relation to risks that cannot be eliminated;
(c) provision of adequate information pursuant to Article 13, in particular as regards the risks referred to in paragraph 2, point (b) of this Article, and, where appropriate, training to users.
In eliminating or reducing risks related to the use of the high-risk AI system, due consideration shall be given to the technical knowledge, experience, education, training to be expected by the user and the environment in which the system is intended to be used.
5. High-risk AI systems shall be tested for the purposes of identifying the most appropriate risk management measures. Testing shall ensure that high-risk AI systems perform consistently for their intended purpose and they are in compliance with the requirements set out in this Chapter.
6. Testing procedures shall be suitable to achieve the intended purpose of the AI system and do not need to go beyond what is necessary to achieve that purpose.
7. The testing of the high-risk AI systems shall be performed, as appropriate, at any point in time throughout the development process, and, in any event, prior to the placing on the market or the putting into service. Testing shall be made against preliminarily defined metrics and probabilistic thresholds that are appropriate to the intended purpose of the high-risk AI system.
8. When implementing the risk management system described in paragraphs 1 to 7, specific consideration shall be given to whether the high-risk AI system is likely to be accessed by or have an impact on children.
9. For credit institutions regulated by Directive 2013/36/EU, the aspects described in paragraphs 1 to 8 shall be part of the risk management procedures established by those institutions pursuant to Article 74 of that Directive».

nell'individuare, valutare e gestire i rischi.

Tale metodologia, impostando la valutazione del rischio in modo flessibile, multifattoriale e proporzionato, appare quindi la "strategia" migliore per garantire la corretta ponderazione dei valori coinvolti nella materia dell'IA, alla luce di quella che – come già evidenziato – è la struttura normativa multilivello esistente nell'ambito della tutela dei diritti fondamentali.

Avvicinando il "rischio" al "fatto" – e allontanandolo dal mero "*legal risk*" – parimenti, si potrebbe garantire una più adeguata risposta anche alle questioni concernenti i profili di responsabilità. Difatti, se il "rischio normativo" lascia spazio solo all'applicabilità di possibili modelli di *corporate criminal liability*[95] – impostati sulla c.d. colpa di organizzazione o sull'applicazione di sanzioni sostanzialmente penali quale mera conseguenza dell'inosservanza degli *standards* richiesti – una concezione multifattoriale e dinamica delle situazioni di rischio consentirebbe di gestire, con le strutture istituzionali penali "tradizionali", quei rischi che tuttora preservino i connotati della ragionevole prevedibilità ed evitabilità[96], in una logica di indagine *case by case*.

7. Come si è evidenziato, la tutela dei diritti fondamentali nel quadro europeo si innesta in un sistema regolatorio "multilivello" e "orizzontale", che impone agli Stati – indipendentemente dalla natura giuridica conferita [97] – l'adozione di sanzioni proporzionate, effettive e dissuasive per la violazione del diritto unionale[98].

L'esigenza di effettività – nel peculiare contesto dell'IA – può ben correlarsi, ad ogni modo, con il principio di residualità dell'intervento penale, come dimostrerebbe l'approccio *risk-based* attuato dall'*AI Act*.

Se è vero che nell'ambito delle attività rischiose, nel corso degli ultimi anni, si è assistito ad un'entropia penalistica, per fronteggiare i rischi "ineliminabili" appare necessario fare appello al principio di *extrema ratio*, che consente di selezionare, tra le

[95] V. Mongillo, *Corporate criminal liability for AI-related crimes: possible techniques and obstacles*, cit., 82 ss.; F. Consulich, *Criminal Law and Artificial Intelligence: Perspective From Italian And Europea Experience*, Munich 2023, 270 ss.

[96] Il grado di colpa andrà, evidentemente, riconnesso al grado del rischio.

[97] Quanto alla nozione di pena adottata dalla Corte EDU, deve anzitutto evidenziarsi che essa fa riferimento alla materia penale di tipo sostanziale individuata – al fine di eludere la c.d. frode delle etichette – con riferimento ai c.d. *Engels criteria.* Cfr. C. eur., GC, 8.6.1976, *Engel e altri c. Paesi Bassi*.

[98] *Ex multis*, si richiama la nota pronuncia della C. eur., 4.3.2014, *Grande Stevens e altri c. Italia*.

varie inosservanze dei consociati, solo quelle davvero intollerabili[99].

In fin dei conti, l'approccio di *risk assessment* proposto dal legislatore europeo – ferme le criticità a cui si è accennato circa la selezione a monte del rapporto tra costi e benefici della commercializzazione e dell'uso di IA completamente autonome – evidenzia la "sopravvivenza" di forme di rischio anche rispetto all'adozione degli *standards*, proponendo un'indicazione – più che di rischio in ordine al verificarsi del fatto di danno – circa il predeterminato "pericolo" di violazione dei diritti fondamentali.

Tale opera di categorizzazione, nonostante la sua limitatezza – atteso che propone un pericolo di danno sulla base di *standard* regolatori individuati a priori senza alcun richiamo al fatto – appare nondimeno essenziale nel definire il parametro oggettivo entro cui determinare la responsabilità per fatto proprio di cui all'art. 27 Cost.[100].

Ne consegue che, nelle aree in cui l'uso di queste IA sia ritenuto troppo rischioso, non così benefico, o non in grado di garantire il dovuto rispetto di alcune prerogative umane fondamentali, i precetti penali dovrebbero, semmai, intervenire *ab origine* al fine: (a) o di regolamentare e all'occorrenza vietare – sul modello dell'*Embryo Protection Act* tedesco[101] – la ricerca e/o l'introduzione nel sistema di questi strumenti "pericolosi"[102]; (b) o di stabilire, comunque, che il mantenimento di un controllo umano significativo debba ritenersi essenziale, garantendo così un centro d'imputazione soggettiva a cui eventualmente addossare la responsabilità[103].

D'altra parte, nelle aree in cui tali tecnologie altamente autonome siano ritenute accettabili, anche a fronte del rischio per la lesione dei diritti – in una prospettiva di prevalente beneficio collettivo – si dovrebbe verosimilmente riconoscere che all'interno dell'area di "rischio socialmente consentito" per ciascun sistema, il diritto penale non rinviene il suo campo applicativo. La tutela dei diritti fondamentali, in tali ambiti – nel pieno rispetto dei canoni di effettività e di sussidiarietà – andrà ricercata nei meccanismi tipici della responsabilità civile, tra i quali, primariamente, la

[99] F. Consulich, voce *Rischio consentito*, cit., 1104.
[100] Si veda, sulla distinzione tra tipicità oggettiva e imputazione oggettiva, M. Cancio Melià, *Líneas Básicas de la Teoría de la imputación objetiva*, Mendoza 2001, 47 ss.
[101] L'*Embryo Protection Act* ("E*mbryonenschutzgesetz*", EschG) dal 1990 regola l'inseminazione artificiale in Germania.
[102] Cfr. K. Gaede, *Künstliche Intelligenz - Rechte und Strafen für Roboter? Plädoyer für eine Regulierung künstlicher Intelligenz jenseits ihrer reinen Anwendung*, Baden-Baden 2019, 76 ss. e 81 ss.
[103] M.E. Florio*, Il dibattito sulla responsabilità penale diretta delle IA: "molto rumore per nulla"?*, cit., 17.

responsabilità da prodotto difettoso[104].

Diversamente opinando, si realizzerebbero forme di responsabilità oggettiva vietata, che rischierebbero, nondimeno, di assumere le vesti di un illecito di mera disobbedienza[105]. Tale modalità di intervento, invero, non appare sufficiente per giustificare un intervento penale[106], essendo ben possibile che una siffatta trasgressione sia talmente distante dal bene giuridico tutelato da violare il principio di necessaria offensività[107]. Difatti, considerato che l'attitudine offensiva di una condotta costituisce il «presupposto dell'ammissibilità della scelta di tutela preventiva sotto il profilo della ragionevolezza, si comprende come risulti estremamente problematico rintracciare uno spazio di legittimazione a fattispecie penali edificate sul principio di precauzione, che rispetto al percorso comunemente seguito nella verifica delle premesse di legittimazione dell'intervento punitivo, e dello stesso modello del pericolo astratto, sembra proporre un vero e proprio "ribaltamento epistemologico", assumendo l'incertezza scientifica sulla pericolosità di talune condotte o situazioni come *starting point* per il relativo divieto penale»[108].

Il principio di precauzione si presenta quindi – come ribadito da autorevole dottrina[109] – come una possibile risposta alla c.d. "incertezza del rischio", ossia come un criterio per adottare misure di protezione prima che si raggiunga una prova scientifica circa il pericolo di danno di una data attività[110]. Ne deriva tuttavia che, specie nel contesto del "rischio da intelligenza artificiale", la decostruzione del *legal risk europeo* nel contesto normativo penale, debba necessariamente prendere le mosse sia

[104] F. Consulich, *Criminal Law and Artificial Intelligence: Perspective From Italian And European Experience*, cit., 291 ss.

[105] Tale ricostruzione potrebbe piuttosto essere compatibile con una "rinnovata" responsabilità penale degli enti, come proposto da V. Mongillo, *Corporate criminal liability for AI-related crimes: possible techniques and obstacles*, cit., 82 ss.

[106] E. Corn, *Il principio di precauzione nel diritto penale. Studio sui limiti all'anticipazione della tutela penale*, Torino 2013, 95.

[107] A. Massaro, *Principio di precauzione e diritto penale:* nihil novi sub sole?, in *www.penalecontemporaneo.it*, 9.5.2011, 13.

[108] V. Manes, *Il principio di offensività nel diritto penale. Canone di politica criminale, criterio ermeneutico, parametro di ragionevolezza*, Torino 2005, 296.

[109] Per una completa disamina dei rapporti tra colpa e rischio nel contesto dell'intelligenza artificiale, si rinvia a L. D'Amico, *Colpa, precauzione e rischio. Le tensioni penalistiche nella moderna era tecnologica*, in *www.lalegislazionepenale.eu*, 21.10.2023. L'autrice evidenzia, infatti, come la colpa, nel settore dell'intelligenza artificiale, sia un capitolo in evoluzione. Con riferimento a tale ambito, si potrebbe dunque pensare di non intendere più la colpa come un'imputazione per la causazione di un evento bensì come un rimprovero per la perdita del controllo (*rectius*, per l'omesso controllo) di una fonte di rischio.

[110] G. Forti, *La "chiara luce della verità" e "l'ignoranza del pericolo", Riflessioni penalistiche sul principio di precauzione*, in AA.VV., *Scritti per Federico Stella*, Napoli 2007, 601.

da una valutazione proporzionale dei "pericoli" per la lesione dei diritti fondamentali che dall'applicazione della categoria del rischio consentito[111].

8. Conclusivamente, può osservarsi come il legislatore europeo, con l'adozione dell'*AI Act,* abbia assunto un ruolo centrale nell'ambito della definizione di un nuovo "nucleo di valori costituzionali" quanto alla tutela dei diritti nel contesto dell'intelligenza artificiale.

Tale assetto regolatorio – pur limitato e auspicabilmente rivedibile nell'ottica del processo di avvicinamento del "rischio legale da intelligenza artificiale" al fatto – pone l'interprete e il penalista di fronte a un nuovo metodo di indagine della responsabilità "indiretta" dell'uomo rispetto agli scenari lesivi derivanti dall'uso dalle macchine autonome, orientato, specificamente, ad un *risk assessment* di "tipo normativo".

Difatti, l'*AIA* mostra come le sfide poste dall'intelligenza artificiale debbano tenere conto di un rapporto tra rischi e benefici collettivi, dove è l'uomo – e non la macchina – ad effettuare, a monte, il giudizio di bilanciamento multifattoriale tra i complessi "valori in gioco"[112]. In questa logica, se, da un lato, pare ribadirsi la centralità del controllo umano e l'"inutilità" teorica di ricorrere a meccanismi di responsabilità "diretta" del sistema di IA[113], dall'altro, appare opportuna – in una logica *risk-based* rispettosa dei canoni di responsabilità personale colpevole *ex* art. 27 Cost. – l'individuazione di aree di c.d. "rischio consentito". Tale approccio, peraltro, non appare nuovo nel settore penale, trovando già applicazione in alcuni ordinamenti europei, quale quello svizzero[114].

---

[111] G. Forti, *La "chiara luce della verità" e "l'ignoranza del pericolo"*, cit., 589 s. Difatti, come osservato anche da F. Consulich, Flash offenders. *Le prospettive di* accountability *penale nel contrasto alle intelligenze artificiali deviant*, in *RIDPP*, 3, 2022, 1053 ss., nei contesti tecnologici l'evento avverso può essere spesso una variabile indipendente dall'implementazione della cautela, giungendo così alla creazione di rischi illeciti anche quando siano rispettate tutte le misure di sicurezza.

[112] Come osservato da L. Stortoni, *Angoscia tecnologica ed esorcismo penale*, in *RIDPP* 2004, 83, infatti, di fronte alle nuove sfide poste dall'innovazione tecnologica, potenzialmente foriere di pericoli, il legislatore deve infatti compiere una scelta di tipo "pre-penale", volta cioè a decidere se consentire lo svolgimento di una data attività oppure proibirlo a monte.

[113] M.E. Florio, *Il dibattito sulla responsabilità penale diretta delle IA: "molto rumore per nulla"?*, cit., 14 ss.

[114] In Svizzera, pur non esistendo un'espressa codificazione nel codice penale, questa teoria ammette esplicitamente che alcuni rischi siano accettati nel diritto penale. Essa consiste nel sostenere che l'esistenza di un rischio è intrinsecamente legata a ogni attività umana e che, di conseguenza, un rischio ragionevole deve essere consentito dal diritto repressivo, anche se si concretizza causando un danno a terzi. Si tratta quindi di una teoria originale che cerca di stabilire in quali condizioni il rischio debba essere consentito. Questa teoria è stata oggetto di numerosi approfondimenti dottrinali, principalmente da parte degli autori di lingua tedesca, ed è stata addirittura recepita, in linea di principio, dalla giurisprudenza del Tribunale federale svizzera. Si rimanda,

Il potenziale intervento del diritto penale in materia, laddove ritenuto essenziale, dovrebbe comunque restare permanentemente incentrato su responsabilità umane[115] o su modelli applicativi di una "rinnovata" [116] *corporate criminal liability* [117], sostanziandosi principalmente nella colpa c.d. di organizzazione o nell'applicazione di sanzioni correlate al mero disallineamento dalle regole di condotta preventive stabilite *ex ante* dall'*AIA* per quella categoria di sistemi[118]. Altrimenti, come proposto da recente dottrina[119], si dovrebbe intervenire nel contesto della deterrenza, applicando sanzioni sostanzialmente penali correlate all'intenzionalità dell'azione perpetrata dall'IA; parimenti, le sanzioni di diritto civile potrebbero ricadere sugli utenti sulla base della loro scelta di utilizzare il sistema di IA e indipendentemente dalla loro colpa. Difatti, come si è osservato, «se le categorie penalistiche non possono essere slabbrate per

per ulteriori approfondimenti, a S. Maeder, M.A. Niggli, «Art. 12», in *Basler Kommentar*[4], a cura di M.A. Niggli e H. Wiprächtiger, Basel 2019, 198: «*Zahlreiche Tätigkeiten sind mit (generell) durchaus vorhersehbaren Gefahren für Rechtsgüter anderer verbunden, die sich nur um den Preis ausschliessen liessen, dass man das entsprechende Verhalten gänzlich untersagte*». Quanto alle pronunce giurisprudenziali, cfr., in particolare, le decisioni del Tribunale federale svizzero ATF 90 IV 8, ATF 116 IV 306, ATF 117 IV 58, ATF 134 IV 193.

[115] In tale logica "umano-centrica" si muove anche il legislatore italiano con lo schema del disegno di legge recante disposizioni e delega al governo in materia di intelligenza artificiale, approvato il 23.4.2024. In particolare, in materia penale – al capo V, artt. 25 ss. – si prevedono alcune ipotesi aggravanti per reati commessi tramite IA, oltre all'introduzione della fattispecie *ex* art. 612-*quater* Cp in tema di illecita diffusione di contenuti generati o manipolati con sistemi di intelligenza artificiale.

[116] Il modello attualmente esistente di responsabilità penale delle persone giuridiche, come è noto, è difatti correlato alla commissione di un illecito penale, in tutti i suoi elementi costitutivi (*actus reus* e *mens rea*) da parte di un agente umano. Attesa l'impossibilità di considerare l'IA alla stregua di un "dipendente" dell'ente, *de facto*, appare inattuabile il tradizionale modello di *corporate criminal liability*. Alcuni autori, nonostante ciò, hanno prospettato tale soluzione. Vedasi, in particolare, M.E. Diamantis, *The Extended Corporate Mind: When Corporations Use AI to Break the Law*, in *North Carolina Law Review,* 2020, 893.

[117] A.F. Tripodi, *Uomo,* Societas, Machina, in *www.lalegislazionepenale.eu*, 10.5.2023, 12 ss.; V. Mongillo, *Corporate criminal liability for AI-related crimes: possible techniques and obstacles*, cit., 82 ss.

[118] Quest'ultima impostazione, secondo V. Mongillo, *Corporate criminal liability for AI-related crimes: possible techniques and obstacles*, cit., 82 ss., appare quella di maggiore compromesso tra le esigenze di prevenzione e l'attuazione della responsabilità penale colpevole. Secondo l'autore, l'attuazione di un modello come quello della colpa di organizzazione, difatti, rischierebbe, in un contesto di rischio "imprevedibile", di addossare all'ente qualsiasi tipo di rischio, in attrito con il principio di responsabilità penale colpevole. Inoltre, allo stato dell'arte, l'ascrizione di una responsabilità corporativa "diretta" dell'IA, previo riconoscimento di una sua personalità giuridica, appare difficilmente prospettabile, attesa la mancata riscontrabilità in capo a tali strumenti dei connotati del "libero arbitrio" e dell'"autocoscienza".

[119] In tema di c.d. *hard crimes*, vedasi il contributo di E. Nerantzi, G. Sartor, *'Hard AI crimes': The Deterrence Turn*, Oxford 2024, 1–29. Il "paradigma di deterrenza dell'IA", secondo gli autori, contribuirebbe a concretizzare il "dovere di diligenza" che i committenti dovrebbero esibire al momento dell'impiego di un agente di IA con le caratteristiche tecniche di una macchina economica. In particolare, i distributori di agenti di IA, da parte loro, sarebbero tenuti a mantenere un meccanismo di conformità (nel caso in cui le tecnologie appropriate siano disponibili nei domini applicativi pertinenti), istituendo (i) *machinae economicissimae*, che includono le sanzioni attese nel loro calcolo di utilità; (ii) *machinae legales*, dotate di un modulo di conformità imperativo; oppure (iii) *machinae legales et economicae*, che combinano i due approcci.

imputare eventi imponderabili, ci si deve limitare ad aggredire, con pene proporzionate», il «disvalore di condotte» che possono costituire forme «di *miscompliance* individuale rispetto a protocolli preventivi nell'impiego di strumenti tecnologicamente avanzati»[120].

In questa prospettiva, la tutela giuridica nel campo dell'intelligenza artificiale – in linea con la giurisprudenza sovranazionale[121] – riaffermerebbe a pieno titolo il ruolo di *extrema ratio* proprio della *matière penale*. Quest'ultima potrà porsi quale soluzione residuale rispetto agli strumenti forniti da altre branche del diritto, pur idonei ad assicurare, in un contesto di "rischio socialmente utile", l'effettività della tutela dei diritti fondamentali dell'Unione[122].

Quello che pare indubbio, in definitiva, è che l'*AI Act* segna una vera e propria "palingenesi valoriale"[123] in materia di intelligenza artificiale; il bilanciamento tra interessi e la gestione "normativa" del rischio (o, forse – sarebbe meglio dire – del predeterminato "pericolo" per la lesione dei diritti), invero, pone il diritto penale nelle condizioni di recuperare la sua vera "umanità".

[120] F. Consulich, Flash Offenders. *Le prospettive di* accountability *penale nel contrasto alle intelligenze artificiali devianti*, cit., 1053 ss.
[121] D. Pranka, *The price of medical negligence – Should it be judged by the criminal court in the context of the jurisprudence of the European Court of Human Rights*?, cit., 124 ss. Vedasi, tra le altre pronunce, C. eur., 11.7.2017, *Mardosai v. Lithuania*.
[122] In tal senso, la Risoluzione del Parlamento europeo del 16.2.2017 recante raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica (2015/2103(INL)) ha suggerito la possibilità di istituire uno "schema di assicurazione obbligatoria" per danni di questo tipo, come pure un fondo teso a coprire i danni non coperti da polizza assicurativa.
[123] L. Picotti, *Traditional Criminal Law Categories and AI: Crisis or Palingenesis*?, in *Traditional Criminal Law Categories and AI: Crisis or Palingenesis?*, a cura di L. Picotti e B. Panattoni, cit., 11 ss.