GIURISPRUDENZA PENALE

Per un uso responsabile dell'IA nel processo (penale)

di ***Andrea Craviotto***, ***Lorenzo Vitali***, ***Luca Egitto*** e ***Chiara Canepa***

L'Intelligenza Artificiale (IA) -da qualche anno- ormai si preannuncia come una nuova innovazione "dirompente" guidata dalla tecnologia, capace di trasformare le nostre vite, il lavoro, le relazioni sociali e i modelli di business consolidati. I recenti e rapidi progressi, soprattutto nel campo dell'Intelligenza Artificiale cd. generativa, hanno suscitato ampi dibattiti sulle molteplici possibilità di applicazione pratica, e sul necessario governo del fenomeno in settori molto sensibili, come la gestione delle infrastrutture critiche, la sanità e, ovviamente, l'amministrazione della giustizia[1].
In questo contributo (a più "voci"), si descrive uno scenario d'uso dell'IA limitato allo specifico ambito del processo penale, che può avere un'applicabilità immediata. E nella descrizione di tale scenario, pur prendendo in esame gli interrogativi e le possibili criticità legate all'utilizzo della IA, si proverà a sgombrare il campo da alcune suggestioni e timori, più o meno fondati, quali, ad esempio, la possibilità che lo sviluppo della IA sfugga al controllo umano (un po' come il computer di bordo HAL 9000 di Odissea nello Spazio), fino ad arrivare a scenari apocalittici causati dalla possibilità teorica che si verifichino eventi catastrofici provocati dalla imprevedibilità degli algoritmi e dalla conseguente autonomia delle macchine da qualsiasi controllo umano. Tutti i punti di vista devono ovviamente trovare spazio nel dibattito pubblico, come ha recentemente ricordato anche il nostro Presidente Mattarella nel suo ultimo discorso di fine anno, ma riteniamo utile e auspicabile che, in parallelo, si trovino ambiti di applicazione immediati su cui iniziare a valutare le potenzialità di queste tecnologie, in modo severamente controllato e responsabile. Ad esempio, è del tutto evidente che debba essere bandita qualsiasi iniziativa volta a sostituire il libero convincimento del giudice e/o il libero apprezzamento delle prove assunte nel processo, poiché ciò minerebbe irrimediabilmente e alla radice l'ordinamento processuale. Di contro, non si comprende per quale ragione un utilizzo mirato, e teso a semplificare attività per loro natura estremamente dispendiose in termini di tempo, come la catalogazione e il caricamento su infrastrutture informatiche degli atti processuali, o la possibilità di compiere verifiche incrociate su fonti di prova tra loro

[1] L "Artificial Intelligence Act" su cui Commissione Europea, Parlamento e Consiglio hanno raggiunto l'intesa politica ha incluso tra i sistemi a più alto rischio quelli progettati per "*…coadiuvare l'autorità giudiziaria nella ricerca e nell'interpretazione dei fatti e della legge, e nell'applicazione della legge a un insieme concreto di fatti*"

eterogenee, non possa essere ammesso (se non addirittura incentivato), affinché le valutazioni che poi ciascun protagonista del processo potrà compiere nell'esercizio del proprio ruolo, siano il più possibile aderenti alla cd. "verità processuale", che è il bene giuridico primario del processo.
In altre parole, uno scenario d'uso auspicabile (già oggi), è quello di colmare il gap attualmente esistente, a livello di efficienza gestionale, tra modalità analogiche/manuali e modalità digitali nel "governo" dei processi penali di grosse dimensioni e di lunga durata nei vari ambiti (da quelli in materia di criminalità organizzata, a quelli in materia di criminalità economica, ove il reperimento delle fonti di prova produce una mole di "carta" impressionante fin dalla chiusura delle indagini preliminari), tale da permettere a tutti gli operatori coinvolti di conservare il pieno controllo del fascicolo fino al termine del processo.
A chiunque abbia svolto la pratica forense, soprattutto nei decenni passati, quando davvero la "carta" dominava ogni aspetto della vita professionale dell'avvocato (e non solo), suonerà familiare la situazione-tipo qui di seguito descritta. Otto di sera, stanza dei praticanti, la giornata lavorativa volge al termine, quando irrompe in stanza il *dominus* di studio: "*Scusate, domani viene sentito il teste Tizio, mi devo preparare le domande per il controesame, mi potete dire se i testimoni già sentiti nelle scorse udienze, hanno detto qualcosa sulla presenza di Tizio alla riunione che c'era stata in azienda o se nelle e-mail sequestrate, Tizio risulta tra i destinatari?*". Panico generale: nessuno dei praticanti ovviamente ha la risposta pronta, ma il *dominus* vuole la risposta, e quindi tocca al praticante più giovane mettersi a scartabellare tra i vari faldoni. Quando finalmente (!), il praticante riesce a trovare le trascrizioni utili e le *e-mail* sequestrate, si accorge, a malincuore, che per poter rispondere a tono alla domanda, dovrà fare le ore piccole in studio per poter passare in rassegna centinaia e centinaia di pagine…
Ora, è del tutto evidente che, nella situazione appena descritta, il praticante "in carne e ossa" non svolge alcun lavoro creativo o valutativo, ma si limita a ricercare e reperire una certa informazione che gli è stata richiesta, ma che, non essendo "esposta" come la merce nelle corsie del supermercato, per poter essere reperita, richiede lo scrutinio di molte fonti informative, spesso eterogenee tra loro e "sparpagliate" tra gli atti del processo (pensiamo al caso in cui vi siano anche intercettazioni telefoniche o ambientali). Ebbene: si può pensare alla messa in opera di un applicativo informatico, che, con l'utilizzo della IA in una particolare declinazione (e che verrà poi illustrata in seguito), vada alla ricerca delle informazioni che servono all'avvocato per orientare al meglio la difesa del proprio assistito. E se gli volessimo dare un nome, potremmo definirlo un "praticante virtuale", senza offesa ovviamente per i praticanti veri, preziosa risorsa di ogni studio legale.
Il diritto penale è il diritto del "fatto", volendo estremizzare potremmo anche dire che, nel diritto penale, è il "fatto" che "crea" il diritto. Ecco, quindi, che la

ricostruzione esatta dei fatti che attengono alla contestazione, permette al difensore di "dare un volto" alla vicenda processuale, e quindi di poter (fondatamente) sostenere una determinata tesi piuttosto che un'altra. Ma, per poter "dare un volto" al fatto, occorre reperire negli atti il maggior numero di elementi che ne consentano un "identikit" preciso e poi sottomettere tali elementi alla semantica non solo del lessico giuridico (dove i termini non hanno lo stesso significato che hanno nel linguaggio comune), ma anche alla particolare semantica espressa dalla specifica vicenda processuale, che avrà connotazioni sue proprie, seppur ascrivibili a categorie concettuali generali (se si tratta di un processo di bancarotta, l'operazione distrattiva o illecita cambia di volta in volta a seconda dei casi, ma i principi generali a cui, per esempio, in questa materia è legata la responsabilità penale degli amministratori senza delega, sono consolidati).

Ed è precisamente in questo ambito che si dovrebbe collocare l'uso "guidato" dell'Intelligenza Artificiale, non già lasciata libera di dare sfogo al proprio fermento "creativo", cosa che inquinerebbe e minerebbe l'affidabilità nella risposta all'interrogazione posta, ma confinata solo nel mondo "chiuso" degli atti del processo, con l'obiettivo di sottoporli a uno scrutinio severo per trarre da essi informazioni univoche e non opinabili, in grado quindi di "unire i puntini" oggettivamente esistenti.

Ciò detto, è noto che un sistema di Intelligenza Artificiale (IA) è costituito da un software che, sotto controllo, riesce a replicare, almeno parzialmente, alcuni meccanismi cognitivi umani. Con la IA il calcolatore non è più solo una scatola nera in grado di eseguire velocemente dei calcoli seguendo algoritmi predeterminati, operando dunque in ottica prettamente procedurale, ma diventa un agente digitale che, dopo una fase iniziale di apprendimento, è in grado di comprendere il contesto nel quale opera, di agire su di esso e di migliorarsi sulla base dell'esperienza e della supervisione umana. Una delle conseguenze di questo cambio di paradigma, è che un algoritmo basato sull'IA produce risultati non sempre prevedibili a-priori, talvolta (addirittura) contrari alle attese, risultando però più efficace in attività come l'interpretazione di suoni/voci (es. trascrivere una conversazione), di un'immagine (es. riconoscere l'identità di un soggetto o il contesto nel quale una foto è stata scattata) o di un testo (es. riassumerlo, tradurlo in altra lingua, interpretarne il senso).

A questo punto possiamo identificare tre famiglie di tecnologie di IA utilizzabili per l'analisi degli atti del processo nell'ottica del "praticante virtuale" a cui si faceva cenno prima[2].

[2] la lista è funzionale al caso descritto ed è dunque incompleta, non includendo, ad esempio, le soluzioni utilizzabili per scopi puramente investigativi, per la stesura dei testi, per la consultazione delle norme, etc.

Alla prima famiglia appartengono gli algoritmi capaci di estrarre l'informazione dagli atti del processo, qualora non immediatamente fruibile. Ad esempio, se un documento in PDF contiene delle immagini, queste devono essere preventivamente tradotte in testo per poter essere utilizzate da un algoritmo. Analogamente, è necessario trascrivere un audio prima di poter trattare in modo automatico il suo contenuto. Ad oggi esistono molteplici tecnologie in grado di automatizzare queste fasi, ed è ragionevole attendersi che in futuro sarà possibile rendere fruibile in modo automatico la totalità di informazioni contenute negli atti del processo.
Alla seconda famiglia appartengono gli algoritmi in grado di estrarre conoscenza dall'informazione digitalizzata, ad esempio per identificare e confrontare il *contenuto* all'interno di testi, immagini e video digitali. Il caso dei testi è direttamente applicabile a un ampio dominio degli atti processuali. Ipotizziamo, ad esempio, di volere identificare, negli atti di un processo, tutti i riferimenti a un determinato trasferimento di denaro, intorno a cui ruota l'intera vicenda processuale. Tale evento sarà citato non solo in atti diversi, ma da fonti diverse e anche in modi diversi, con espressioni come: *pagamento*, *bonifico*, *assegno*, *movimento*, *transazione*, *versamento*, etc. Talvolta sarà citato l'importo, talvolta no, e oltretutto il riferimento a tale evento potrebbe essere del tutto indiretto. Per poter, dunque, collegare i riferimenti allo stesso evento, sarà necessario automatizzare un passaggio concettuale (che l'uomo compie spontaneamente quando legge un testo), astraendo il *concetto* di trasferimento di denaro dal contesto e dalle modalità nelle quali è stato espresso. Le moderne tecniche di IA sono oggi in grado di calcolare in tempo reale la vicinanza tra il *significato* di informazioni digitali "sparse" su un dominio anche molto esteso. Il principio utilizzato da questa tecnologia è la capacità di trasformare una certa informazione, ad esempio lo stralcio di una deposizione resa in una certa udienza, nella rappresentazione numerica del suo significato, una procedura che rende possibile la creazione di una vera e propria "mappa semantica" multidimensionale, nella quale contenuti di significato uguale o simile, si troveranno vicini.
Se la immaginiamo su due dimensioni, sarebbe come collocare le foto di frutti su una mappa geografica, assegnando ad ognuna un diverso vettore "latitudine-longitudine" tale per cui, ad esempio, tutte le pesche saranno tra loro vicine, a poca distanza dal gruppo delle albicocche, ma molto distanti dal gruppo delle banane, e che, a loro volta, saranno distantissime dal gruppo delle olive. E così via, in quanto la reciproca "distanza" sarebbe rappresentativa del contenuto delle foto ossia della forma, della dimensione e del colore dei vari frutti in esse rappresentati. Gli algoritmi di IA sono in grado di realizzare questa mappatura su un numero arbitrario di dati e su migliaia di dimensioni, e di registrarla su appositi database vettoriali che possono essere interrogati in tempo reale. Grazie a queste tecnologie è

dunque ipotizzabile di realizzare un'analisi di vicinanza semantica sulla totalità delle fonti di prova acquisite.
Alla terza categoria appartiene infine la cosiddetta Intelligenza Artificiale Generativa (AIG), abilitata da modelli linguistici di grandi dimensioni (LLM), in grado di generare testi e altri contenuti digitali a partire dal contenuto su cui sono stati addestrati, e sulla base di una specifica richiesta dell'utente. Se i database vettoriali permettono, dunque, di compiere l'operazione fondamentale di scrutinio degli atti, estraendone il *significato* e trovando correlazioni tra quantità corpose di dati, i modelli LLM vanno oltre, essendo in grado di comprenderne il significato semantico più profondo, di sviluppare un "ragionamento" linguistico sotto la richiesta di un essere umano, come riassumere un testo, farlo con un certo numero di battute, adottando un certo stile di scrittura, o addirittura un "tono" (satirico/aulico).
È evidente che l'utilizzo di modelli LLM per l'analisi degli atti di un processo richiede la capacità di comprendere la lingua italiana, le sue sfumature, e in particolare il senso che emerge nelle interazioni domanda-risposta con cui viene generata l'informazione nel corso delle udienze dibattimentali.
A questo proposito riteniamo utile "sfatare" il mito dei potenziali rischi derivanti da un ipotetico processo di apprendimento dei modelli LLM, se addirittura del loro semplice utilizzo, che implicherebbe la memorizzazione di dati confidenziali e la loro rivelazione a terzi che dovessero utilizzare gli stessi modelli successivamente. Come spiegheremo più avanti inquadrando il tema in un contesto più ampio di compliance e sicurezza delle soluzioni di IA, lo scenario ipotizzato non prevede infatti di usare modelli pubblici ma solo delle loro versioni private che scongiurano tale eventualità. Inoltre, grazie all'utilizzo dei database vettoriali che costituiscono una memoria semantica degli atti del processo, la quantità di informazione analizzate dai modelli LLM risulta comunque minimizzata. Allo stato attuale di sviluppo della tecnologia, dunque, il contributo dei modelli LLM è determinante ma si limita alla conoscenza della lingua italiana e alla capacità di ricostruire i nessi causali tra circostanze citate in testi sparsi. Una facoltà che, come detto, non richiede l'apprendimento del motore sugli atti del processo.
Un altro tema che normalmente emerge di fronte a queste soluzioni è l'affidabilità della loro risposta. A questo proposito si devono fare due osservazioni. La prima è che i motori LLM sono stati progettati per bilanciare obiettività e creatività. Chiunque abbia provato le varie soluzioni disponibili in rete, ha certamente osservato che la loro obiettività è gradualmente migliorata. Le prime versioni commettevano errori banali, e talvolta generavano risposte totalmente inventate (fenomeno detto di *allucinazione*), problemi che sono stati gradualmente limitati ma non risolti completamente. La soluzione migliore negli scenari descritti è dunque quella di procedere a una istruzione preventiva della macchina su '*come*' rispondere (non sul '*cosa*'). Questa funzionalità, denominata *prompt engineering*, è una delle nuove

professioni abilitate dalla IA, che permette di istruire l'algoritmo a rispondere solo sui dati di contesto, a non inventare, a non prendere posizione nel caso di affermazioni contraddittorie, a non utilizzare informazioni, notizie e/o commenti provenienti da "fonti esterne" estranee al dominio dei dati sottomesso alla IA. Sarà comunque sempre possibile che l'algoritmo fornisca risposte imprecise, ad esempio nel caso non riesca a interpretare argomentazioni involute o affermazioni contraddittorie, ma ciò non è da ritenersi un problema, dal momento che la finalità del suo utilizzo non è certo quella di sostituirsi alle valutazioni discrezionali umane. Anzi, la validazione umana, nel settore della amministrazione della giustizia, resta un requisito imprescindibile e mai rinunciabile. Semplicemente l'IA può rendere un servizio utile e controllato.

Merita -infine- un'ultima considerazione la giusta preoccupazione relativa alla confidenzialità delle informazioni e la protezione dei dati personali trasmessi ai motori di LLM. Si tratta di una preoccupazione fondata non solo sulla base elementi intuitivi rispetto ai principi generali espressi dal GDPR, ma anche su esplicite osservazioni del legislatore europeo[3] e di numerosi esperti[4].

L'utilizzo di tecnologie innovative nel trattamento dei dati personali è tipicamente uno degli scenari in cui è obbligatorio valutare eventuali impatti negativi per i soggetti interessati derivanti dall'uso di tali tecnologie. L'intelligenza artificiale ricade sicuramente in questi scenari.

I sistemi di intelligenza artificiale hanno infatti caratteristiche tali da prospettare effetti potenzialmente pregiudizievoli per i soggetti interessati, laddove l'uso dei loro dati risulti non trasparente, non proporzionato o altrimenti fuori dal controllo degli interessati stessi. A tali rischi si sommano poi quelli legati alle conseguenze ulteriori del trattamento, da più parti ritenute potenzialmente discriminatorie o addirittura traumatiche.

Il livello di severità in tali ambiti è frutto della combinazione di due rischi. Il primo rischio è inerente al sistema di addestramento ed elaborazione delle informazioni dell'IA, che in quanto tale rappresenta un metodo di trattamento inusuale e spesso poco chiaro persino per coloro che lo programmano. A questo si aggiunge il rischio storicamente collegato alle tipologie di dati trattati. È evidente, infatti, che in ambito giudiziario

[3] si vedano, es., il preambolo 15, 36 e 45 nonché art. 10 (5) della bozza dell'AI Act https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206.

[4] tra i tanti si possono citare: *Emiliano De Cristofaro*, 2020 "An overview of privacy in machine learning" https://arxiv.org/abs/2005.08679; *Veale, Michael, Reuben Binns, and Lilian Edwards*. 2018. "Algorithms That Remember: Model Inversion Attacks and Data Protection Law" Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 376 (2133): 20180083. https://doi.org/10.1098/rsta.2018.0083.

l'esercizio difensivo, dovendo accedere a enormi volumi di dati spesso di categorie particolari, rappresenti un trattamento di dati personali ad alto rischio. Il ricorso a sistemi legal-tech fondati sull'IA – e prima ancora il design di tali sistemi - deve considerare con massima severità e accuratezza i rischi per i diritti fondamentali connessi all'uso di tali tecnologie, valutando sia l'impatto privacy sia quello etico. Tale valutazione deve assicurare che la tecnologia progettata rispetti i principi fondamentali di tutela dei diritti umani e di protezione dei dati personali, in particolare per garantire correttezza, trasparenza, limitazione delle finalità del trattamento, minimizzazione ed esattezza dei dati, limitazione della conservazione e sicurezza del trattamento.

Spostando il focus di tale analisi sul caso in esame, va osservato in primo luogo che lo scenario descritto non prevede alcuna forma di addestramento o "training" dell'algoritmo, ma solo il trasferimento dei dati di contesto al motore LLM, nella forma di un insieme di frammenti sparsi tra gli atti processuali che sono rilevanti per un certo contesto quando si "interroga" il sistema per reperire una determinata informazione contenuta nei dati.

Questo primo discrimine è fondamentale perché esclude una componente critica del trattamento, l'addestramento del sistema di IA attraverso l'uso di dati personali, spesso considerato troppo poco trasparente per essere ritenuto sufficientemente controllabile e senza conseguenze per gli interessati. La finalità del trattamento è quella di individuare con massima precisione informazioni necessarie per esercitare un diritto in sede giudiziaria e tale scopo è delimitato dall'ecosistema dell'LLM strettamente necessario per eseguire e dare seguito a una interrogazione, evitando processi non essenziali.

I dati vengono analizzati in tempo reale dal motore LLM, senza memorizzarli, e senza memorizzare neppure la risposta data. Dovendoci limitare ai motori LLM in grado di comprendere la lingua italiana, ad oggi queste soluzioni sono disponibili in ambienti *cloud*, all'interno della UE, in contesti in cui è garantita la minimizzazione del trattamento con misure di sicurezza ai massimi standard.

Le misure di sicurezza applicate in questo scenario sono una combinazione di presìdi delle informazioni a riposo e in transito (che possono includere crittografia e mascheramento) con misure tecnico-organizzative di difesa contro vulnerabilità e attacchi di avversari ostili. Le misure di sicurezza della tecnologia legata all'IA devono essere coordinate tra le funzioni remotizzate e quelle sotto il controllo diretto dell'utilizzatore. Infatti, oltre a configurare l'ambiente cloud con misure proporzionate ai livelli di rischio più elevati occorre avere un focus analitico sulla sfera dell'avvocato (e dei professionisti delegati dallo stesso) che utilizza il sistema di ricerca "*AI-powered*". La postazione del legale che interroga l'applicativo, è l'ingresso ad una "superficie di attacco" storicamente critica (da molto prima dell'avvento

dell'IA) che richiede di essere presidiata non solo da misure tecnologiche il più possibile avanzate, ma anche dalle misure organizzative tipiche del GDPR che gli studi professionali hanno spesso trascurato e che in questo caso sono ancora più essenziali e ineludibili.
Il tema centrale, oltre la limitazione, proporzionalità e liceità del trattamento, è proprio la sicurezza dello stesso.
L'avvocato difensore, nell'esercizio del mandato ricevuto dal proprio cliente coinvolto in un procedimento giudiziario (penale/civile/amministrativo) tratta dati personali – anche particolari – sotto l'ombrello del diritto all'azione giurisdizionale. Su questa base giuridica, il difensore-titolare del trattamento determina i mezzi con cui perseguire le finalità di difesa ben potendo ricorrere a tecnologie innovative come l'IA per rendere al meglio la propria prestazione professionale nell'interesse (ovviamente) del proprio assistito, laddove il ricorso a tali sistemi non alteri i limiti del trattamento e non introduca rischi ulteriori. La ricerca della prova è strutturalmente connessa al diritto di difesa e l'utilizzo di nuove tecnologie per raggiungere tale prova è non solo plausibile, ma probabilmente necessitato dalla sempre più grande mole di informazioni digitali che il difensore deve consultare per trovare informazioni essenziali alla tutela del proprio cliente. Il punto, quindi, non è "se" sia possibile usare un motore LLM per ricercare efficacemente una stringa di testo tra centinaia di migliaia di pagine, quanto piuttosto "come" usare tali tecniche. È fondamentale, pertanto, che l'avvocato che si avvale di questa tecnologia abbia una *governance privacy* di studio rigorosa, aggiornata e attrezzata ad affrontare le minacce più gravi. Mai come in questo contesto lo studio professionale deve dotarsi di un organigramma e una catena di comando nitida e collaudata, corredata da istruzioni e deleghe di compiti accurate e pertinenti, in cui il reperimento delle informazioni sia organico all'attività difensiva, ma allo stesso tempo inserito in una consapevole e competente gestione delle risorse informatiche e applicazione dei principi di protezione e sicurezza delle informazioni (personali e non).
In sostanza, utilizzare cloud computing e IA è ammissibile in un contesto difensivo come quello descritto, a patto che l'organizzazione e le misure tecniche adottate dai professionisti che utilizzano l'IA nella *data discovery* siano al medesimo livello di robustezza e aggiornamento dei sistemi remotizzati dei grandi *cloud providers*.
In conclusione di questo percorso, ci pare importante dare spazio anche al punto di vista di chi sta dall'altra parte "della barricata", ossia il Pubblico Ministero, parte del processo una volta che questo è instradato nella fase dibattimentale, e che, essendo al tempo stesso il titolare dell'azione penale, ha un ruolo di fondamentale importanza nella fase precedente, quella delle indagini preliminari.
Ospitiamo, quindi, con piacere il contributo della Dr.ssa Chiara Canepa, Sostituto Procuratore della Repubblica presso il Tribunale di Torino.

Il panorama normativo, ancora *in fieri,* vede tra i primi documenti sul tema, la Carta Etica dell'Europa emanata il 4 dicembre 2018 dal Consiglio d'Europa. La Carta Etica si fonda su alcuni principi essenziali, cercando un bilanciamento tra l'uso di sistemi di intelligenza artificiale e la tutela dei diritti fondamentali degli attori del processo. Da viene un lato, viene sottolineata l'importanza di assicurare risultati non discriminatori, dall'altra l'esigenza di preservare l'intangibilità e l'integrità del sistema; ancora, il Legislatore europeo ammonisce l'utente ad utilizzare sistemi in grado di certificare l'autenticità e la correttezza del dato acquisito e l'attendibilità della fonte, Infine, è necessario che tutti i sistemi di IA siano conformi ai canoni di trasparenza (cioè possibilità di accesso ai meccanismi di funzionamento), imparzialità (assenza di "pregiudizi"), equità e integrità intellettuale (si devono, cioè, privilegiare sempre gli interessi pubblici e quelli della giustizia) delle metodologie di trattamento di dati personali sensibili. L'ultimo principio è quello della "verificabilità" del processo di trattamento da parte del fruitore del servizio, in modo da consentirgli di agire in modo informato e di adottare consapevolmente le proprie scelte.
Nell'aprile 2021, la Commissione Europea, al fine di definire regole uniformi sull'utilizzo dei sistemi di IA, ha redatto la *Proposal for a Regulation Of The European Parliament And Of The Council, Laying Down Harmonised Rules On Artificial Intelligence (Artificial Intelligence Act) And Amending Certain Union Legislative Acts*: essa fornisce una definizione di IA, definisce regole di trasparenza e doveri di controllo (ad es. l'obbligo del c.d. conformity assessment) per gli operatori che intendano utilizzare tali tecnologie e, infine, vieta espressamente l'uso di determinati sistemi di IA («prohibited artificial intelligence practices»), che possono facilmente comportare una violazione dei diritti fondamentali. Si prevede, infine, l'istituzione di uno *European Artificial Intelligence Board* quale organo di supporto della Commissione e, allo stesso tempo, di coordinamento per il lavoro delle Autorità di vigilanza a livello nazionale. Tali documenti trovano il loro fondamento non solo nella necessità di regolare le prime sperimentazioni di programmi di cd. "giustizia predittiva" in sede europea, ma anche di affrontare l'utilizzabilità del dato ricavato da processi di elaborazione, condotti da IA, di riconoscimento facciale[5].
Anche l'Italia si è munita di un sistema di riconoscimento facciale nell'ambito dell'attività investigativa e di prevenzione del crimine: dal 2018, l'attività della Polizia di Stato è supportata dal Sistema Automatico di Riconoscimento Facciale (SARI). Il sistema è in grado di operare sia in modalità *Real Time* (ove analizza immagini "live" provenienti da telecamere insistenti su zone di

[5] Cfr. European Union Agency for fundamental rights, Facial recognition technology: fundamental rights considerations in the context of law enforcement, Fra Focus, 27 novembre 2019;

transito ed aree pubbliche) sia in modalità *Enterprise* (quando si confrontano immagini presenti in più banche dati al fine di creare l'*identikit* digitale di maggior precisione).

A parere della scrivente, l'*habitus* del giurista nell'affrontare tali nuovi software deve rivestire il carattere della prudenza[6], senza giungere mai ad un atteggiamento di chiusura. Se, da un lato, il giurista deve poter vagliare la trasparenza ed il cd. *conformity assessment* del programma che sfrutta IA per l'analisi dei dati processuali (essendo posto in grado di conoscere la "semantica" di riferimento del *software* medesimo), dall'altro deve porre in essere una serie di controlli successivi per confrontarsi con il risultato ottenuto, prendendo così coscienza diretta della fallibilità, ma anche delle potenzialità dello strumento *de quo*.

E così, come la paternità del gesto criminoso, attribuito analizzando con il SARI le immagini delle *webcam* del comune di una città sottoposta, ad esempio, ad atti di vandalismo, dovrà essere sostenuta da altri elementi probatori od indiziari a riscontro della "tesi informatica", così il riassunto della deposizione di un teste ottenuto utilizzando un software di IA dovrà essere confrontato con la trascrizione integrale, attività che consentirà, così, all'utente di fruire di una sorta di "brogliaccio" informativo da integrare e correggere nei punti salienti eventualmente omessi o non compresi nel significato più genuino.

Si condivide, in tal senso, l'opinione di illustri esponenti della Dottrina italiana[7] ed estera[8] che auspicano non solo l'implementazione della normativa di principio, con la fissazione di standard minimi di tutela dei diritti fondamentali, ma la redazione a livello comunitario di una disciplina di dettaglio, che garantisca procedure *ad hoc*, anche per la verifica della legittimità nell'utilizzo di tali strumenti, unendo l'esperienza degli informatici con la sensibilità dei giuristi.

---

[6] A. Soro, La protezione dei dati personali nell'era digitale, in Nuova Giur. Civl, 2019, 2, pp. 343 e ss;

[7] https://dirittopenaleuomo.org/wp-content/uploads/2021/05/Currao_DPU.pdf

[8] Vds. C. Garvie, J. Frakle, Facial-Recognition Software Might Have a Racial Bias Problem, in The Atlantic, 7 aprile 2016; R. Morrison, "Racist" facial recognition technology used in law enforcement, banking and schools misidentifies African American and Asian people 100 times more often than whites, study shows, in Dailymail, 19 dicembre 2019.