RASSEGNA ITALIANA DI CRIMINOLOGIA ANNO XVI N.3 2022
Società Italiana di Criminologia
Articoli di ricerca

Psychiatric-forensic evaluations and artificial intelligence: new possible scenario

Valutazioni psichiatriche-forensi e intelligenza artificiale: nuovi scenari possibili

Simona Casale | Stefano Ferracuti | Giovanna Parmegiani

OPEN ACCESS

Double blind peer review

How to cite this article: Casale S. et alii (2022). Psychiatric-forensic evaluations and artificial intelligence: new possible scenario. *Rassegna Italiana di Criminologia*, XVI, 3, 211-219. https://doi.org/10.7347/RIC-032022-p211

Corresponding Author: Simona Casale
email simona.casale@uniroma1.it

Received: 12.04.2022
Accepted: 05.09.2022
Published: 22.12.2022

Pensa MultiMedia
ISSN 1121-1717 (print)
ISSN 2240-8053 (on line)
doi10.7347/RIC-032022-p211

Abstract
Cognitive biases are defined as mental processes which can lead to the elaboration of misjudgements. Biases can influence thoughts, opinions, behaviours, and they are inevitably involved in psychiatric-forensic evaluations. The delicate role that the mental health expert plays in the psychiatric examinations makes these mistakes highly relevant. Make sure that such a complex decision-making process is set up with care and attention is fundamental to guarantee the best protection of the individual rights, but also to avoid a mismanagement of government money. In order to avoid the possible use of cognitive biases in psychiatric-forensic assessments, it has been suggested to increase the standardization of evaluation procedures, but scientific literature shows that experts working in this area are often reluctant to question their own work. Recently, the application of Artificial Intelligence (AI) to the forensic field has opened new possibilities, but, on the other hand, it has generated new questions. AI seems to be beneficial for those decision-making processes where greater standardization is required, but attempts to use AI tools in the field of forensic psychiatry have highlighted some critical issues. In this article, after discussing the problems that characterize both human and computational decision-making, we will propose possible solutions.

Keywords: Artificial Intelligence, Psychiatric-forensic evaluations, cognitive bias, heuristic functions, decision making processes.

Riassunto
Si definiscono "*bias* cognitivi" quei processi mentali che possono condurre a elaborare giudizi imprecisi o errati. I *bias* possono influenzare pensieri, opinioni, condotte e sono inevitabilmente implicati anche nelle valutazioni psichiatriche-forensi. Il delicato ruolo che l'esperto in salute mentale riveste all'interno del processo peritale rende questi errori particolarmente rilevanti. Assicurarsi che un processo decisionale così complesso sia messo in atto con la dovuta attenzione e cura risulta di fondamentale importanza per garantire la maggiore tutela possibile del singolo, ma anche per evitare una mala gestione dei soldi dello Stato. Per sfuggire all'eventuale utilizzo di *bias* cognitivi nelle valutazioni psichiatriche-forensi, si è suggerito di standardizzare maggiormente le procedure di valutazione, ma dalla letteratura emerge che gli esperti che operano in quest'ambito sono spesso restii a mettere in discussione il proprio stesso operato. Di recente, l'applicazione dell'Intelligenza Artificiale (IA) al campo forense ha aperto nuove possibilità, ma, d'altro canto, ha generato nuovi interrogativi. L'IA sembra essere vantaggiosa per quei processi decisionali in cui maggiore standardizzazione è richiesta, ma i tentativi di utilizzo di strumenti dotati di IA in campo forense hanno messo in evidenza alcune criticità. In questo articolo, dopo aver discusso i problemi che caratterizzano sia il processo decisionale umano che quello computazionale, proporremo possibili soluzioni.

Parole chiave: Intelligenza Artificiale, Valutazioni psichiatriche-forensi, bias cognitivi, euristiche, processi decisionali.

Simona Casale, Department of Human Neuroscience, Sapienza University of Rome, Rome, Italy | Stefano Ferracuti, Department of Human Neuroscience, Sapienza University of Rome, Rome, Italy | Giovanna Parmigiani, Department of Human Neuroscience, Sapienza University of Rome, Rome, Italy

Psychiatric-forensic evaluations and artificial intelligence: new possible scenario

Introduzione

Il termine "*bias* cognitivo" indica una varietà di processi che possono portare a giudizi o interpretazioni imprecisi e che possono influenzare la memoria, il ragionamento e il processo decisionale. Il pensiero tende difatti a elaborare scorciatoie cognitive o euristiche al fine di destreggiarsi con maggiore facilità tra una varietà di stimoli complessi (Cooper & Meterko, 2019). I *bias* cognitivi rivestono un ruolo rilevante in tutti gli aspetti della percezione umana e del processo decisionale e, quindi, sono inevitabilmente implicati anche in molti settori della scienza forense e della valutazione forense (Zapf & Dror, 2017). Neal e Grisso (2014) ipotizzano che le euristiche più utilizzate in ambito forense siano l'euristica dell'ancoraggio, della rappresentatività e della disponibilità. Le euristiche, in pratica, aumentano le probabilità che il valutatore tenda a trascurare alcuni elementi che potrebbero essere di fondamentale importanza per la decisione finale, perché assume un atteggiamento poco esplorativo (Skellern, 2015; Zapf & Dror, 2017), divenendo fonte di pregiudizio inconsapevole. L'euristica dell'ancoraggio consiste nel fare una stima partendo da un valore iniziale (àncora) che viene poi aggiustato per ottenere la risposta finale. Il valore iniziale può essere suggerito da come è formulato il problema o può derivare da calcoli incompleti. In entrambi i casi gli aggiustamenti successivi sono tipicamente insufficienti. Punti di partenza diversi ottengono differenti stime finali, che sono distorte dai valori iniziali, che fungono da àncora. L'euristica della disponibilità consiste nella facilità di evocare mentalmente una data situazione: più essa è disponibile per il soggetto, più tende a sembrare probabile. L'euristica della rappresentatività consiste nella tendenza ad associare qualcuno o qualcosa ad una certa classe in base a quanto è rappresentativo di quella classe. Il grado di rappresentatività non dipende da valutazioni statistiche, al contrario, in questo tipo di ragionamento vengono trascurate altre variabili statisticamente rilevanti, come la probabilità di base.

Nel processo decisionale del valutatore, può capitare che un solo elemento rivesta un ruolo così importante da essere in grado di ribaltare il giudizio complessivo, senza considerare l'insieme d'informazioni di cui si dispone (Skellern, 2015), oppure che alcuni elementi che inizialmente apparivano irrilevanti finiscano per assumere un ruolo decisivo (Skellern, 2015; Zapf & Dror, 2017). Il delicato ruolo che l'esperto in salute mentale riveste all'interno del processo peritale rende questi errori particolarmente rilevanti. Nonostante ciò, sembra che la percezione del problema in ambito professionale forense desti scarso interesse, venga sottovalutato e non avvertito come importante o urgente (Gowensmith & McCallum, 2019; Kukucka et al., 2017). Per esempio, da un'indagine condotta su 403 esperti di medicina legale emergeva una generale riluttanza a riconoscere l'importanza di avvalersi di procedure volte a minimizzare pregiudizi cognitivi e una certa resistenza nel riconoscere la propria suscettibilità ai pregiudizi (Kukucka, et al., 2017). Zapf e Dror (2017) evidenziano che il cervello umano ha una capacità limitata di rappresentare ed elaborare tutti i dati e così privilegia l'utilizzo di tecniche volte a sviluppare in modo efficiente e rapido l'informazione. Gli autori riconoscono che, considerata la natura stessa dell'essere umano, è impossibile eliminare tutti i *bias* e che si può solo cercare di minimizzarli quanto più possibile. Di recente, però, l'applicazione dell'Intelligenza Artificiale (IA) in campo forense ha aperto la possibilità di un processo decisionale privo di *bias*, perché, appunto, non più frutto del ragionamento umano. I primi esperimenti fatti in tal senso hanno portato allo sviluppo di nuovi campi d'indagine e nuove domande (Corbett-Davies & Goel, 2018; Kehl & Kessler, 2017). Di seguito andremo a descrivere alcuni *bias* cognitivi che determinano errori procedurali nelle valutazioni psichiatriche-forensi, le soluzioni che sono state proposte dai ricercatori che si sono interessati al tema e il ruolo che l'IA ha e potrebbe avere in quest'ambito.

I *bias* cognitivi nelle valutazioni psichiatriche-forensi

I *bias* sono delle rappresentazioni distorte di alcuni aspetti della realtà oggettiva (Haselton et al., 2015). Un processo decisionale influenzato dai *bias* può causare ingiustizie (Curley et al., 2022). I *bias* non hanno a che fare con corruzione, malizia o incompetenza dell'esperto (un esperto molto competente può essere comunque vittima di *bias*). Difatti, gli esperti in un certo settore non sono meno soggetti a *bias* rispetto ai non esperti e non basta prendere coscienza dell'esistenza dei *bias* per diventare immuni ad essi (Dror, 2020). A differenza di quello che si crede, inoltre, è stato dimostrato che le decisioni prese non si basano sulla complessa combinazione e ponderazione di un gran numero di variabili, bensì sull'analisi di pochissimi aspetti, anche se ci sono delle eccezioni (Arkes, 1989).

Le cause dei *bias* nelle valutazioni forensi possono essere classificate in tre categorie: fattori legati alla natura stessa dell'essere umano; fattori legati alle caratteristiche individuali dell'esperto; fattori legati al caso specifico che l'esperto viene chiamato a valutare (Dror, 2020).

Fattori legati alla natura stessa dell'essere umano. Il cervello umano si impegna in una varietà di processi per dare un senso al mondo e crea vincoli che non consentono

di elaborare tutte le informazioni che riceve (Dror, 2020). L'Homo Sapiens, infatti, ha una capacità cognitiva limitata e tende a prediligere l'efficienza nel prendere decisioni rispetto all'accuratezza (Curley et al., 2022).

Fattori legati alle caratteristiche individuali dell'esperto. I valori dell'esperto, le sue esperienze di vita, le sue idee politiche, la cultura di appartenenza giocano un ruolo rilevante nelle valutazioni (Zapf & Dror, 2017). Ad esempio, in un caso in cui si dibatte in merito alla pena di morte, il professionista che è contrario a questo tipo di pena può essere condizionato dalla sua idea politica durante il procedimento (Goldyne, 2017).

Fattori legati al caso specifico che l'esperto viene chiamato a valutare. Una prima possibile causa di errori nelle valutazioni forensi è proprio la scelta dell'esperto che viene chiamato ad esprimersi in merito al caso in esame. Skellern (2015) definisce questo fenomeno "effetto selezione", riferendosi al fatto che viene scelto un certo esperto piuttosto che un altro perché lo si reputa più adeguato a sostenere le proprie convinzioni rispetto al caso in esame. Questo effetto sarebbe strettamente connesso all' "effetto fedeltà" ovvero alla perdita di oggettività che si verifica quando la valutazione dell'esperto è commissionata da una parte (Gowensmith & McCallum, 2019; Skellern, 2015).

Nominato il professionista, sarà suo compito esaminare la documentazione disponibile. In questa fase, i dati che devono essere analizzati possono essere mal interpretati oppure l'esperto può essere condizionato dal parere di altri nella valutazione di essi (Arkes, 1989; Skellern, 2015).

Per decidere come procedere ai fini della valutazione, inoltre, l'esperto potrebbe essere condizionato da quelle informazioni che vengono definite "contestuali". Si pensi, ad esempio, ad un caso in cui viene richiesta un'autopsia ad un medico legale con un'informazione contestuale relativa al fatto che il defunto aveva una storia pregressa di consumo di eroina. Questa informazione può influenzare la scelta dei test a cui il medico legale sottoporrà il defunto, tralasciando test standard che, se non avesse ricevuto quella informazione contestuale, avrebbe utilizzato (Dror, 2020). Il professionista potrebbe anche tralasciare delle informazioni fondamentali per un processo decisionale corretto. Un aspetto che dovrebbe considerare, ad esempio, è quello dei "tassi di riferimento" cioè la prevalenza di una certa caratteristica nella popolazione di riferimento. Infatti, bisognerebbe considerare quanto, sulla base delle statistiche, quell'ipotesi o quel dato di cui disponiamo nel caso in esame risulti essere credibile. Questa informazione, invece, troppo spesso appare non rilevante agli occhi dell'esperto perché non risulta immediata la relazione causale tra l'evento in questione e i tassi di riferimento (Arkes, 1989). D'altra parte, più comunemente gli esperti si rifanno alla propria esperienza personale nel valutare il caso in esame (Dror, 2020). Quindi se, ad esempio, nella loro esperienza personale, la maggior parte delle volte in cui la persona era incapace di intendere e di volere aveva subito violenze in passato allora tendono ad affidarsi a questo schema, piuttosto che a dati statisticamente rilevanti. Ad inficiare il processo decisionale è anche la memoria, per esempio si tende a ricordare più facilmente le informazioni che confermano la propria opinione (Arkes, 1989). Altro errore è legato al rapporto di causalità che l'esperto "percepisce" tra alcuni fattori relativi al caso in esame, senza di fatto esplorare tutte le possibilità. Questo spesso avviene, perché si tende a dare molto peso alle prime impressioni e a cercare aspetti che le confermino, piuttosto che il contrario (Arkes, 1989).

Sattar et al. (2002) fanno presente il ruolo centrale che può avere il controtransfert anche nelle valutazioni psichiatriche-forensi e la difficoltà a controllare questo fenomeno soprattutto tra i valutatori più giovani e inesperti. Everson e Sandoval (2011) rappresentano che, in generale, i professionisti più giovani o di sesso femminile tendono a sopravvalutare la possibilità che una denuncia di abuso sessuale sia verosimile, mentre i più anziani tendono a sottovalutare questa possibilità. Questi diversi atteggiamenti verso i casi di abuso sessuale in genere aumentano il rischio che il processo decisionale del caso specifico non sia oggettivo e che gli esperti vadano alla ricerca degli elementi che confermano la loro teoria di base.

Quelli finora descritti sono gli errori che Arkes (1989) definisce "freddi" e che distingue dagli errori "caldi", ovvero quegli errori che sono motivati da differenze di razza, genere e livello socioculturale. Mikton e Grounds (2007) trovarono che gli psichiatri forensi, nel Regno Unito, tendevano a sovrastimare la presenza di un disturbo di personalità antisociale tra i caucasici e a sottostimarlo tra i non caucasici, mentre non si verificava la stessa cosa per il disturbo borderline. Notarono, inoltre, che i valutatori di altre etnie avevano interiorizzato gli "standard caucasici" e le loro diagnosi non erano diverse in modo significativo da quelle degli altri colleghi. Da vari studi è emerso che i valutatori, per lo più caucasici, tendono a fare diagnosi di disturbo antisociale e paranoide agli afroamericani, di disturbo schizoide agli asiatici e di disturbo schizotipico agli americani indiani, mentre gli altri disturbi di personalità vengono più frequentemente attribuiti ai caucasici. Ricerche dimostrano anche che di solito gli esperti sovrastimano la possibilità che un afroamericano sia stato violento, mentre per i caucasici questa possibilità è sottostimata (Hicks, 2004). Vari studi, inoltre, evidenziano che gli afroamericani ricevono più spesso diagnosi di disturbo psicotico (Blow et al. 2004; Neighbors et al. 1999; Strakowski et al. 2003). Perry et al. (2013) avevano ipotizzato che la tendenza a diagnosticare più facilmente il disturbo psicotico agli afroamericani, poteva agevolarli durante il procedimento penale perché chi riceve questo tipo di diagnosi più frequentemente viene reputato incapace di intendere e di volere. Gli autori analizzarono 129 perizie psichiatriche e verificarono il ruolo che diverse variabili giocavano sulla decisione finale. Trovarono che, qualora l'imputato avesse ricevuto una diagnosi di disturbo psicotico, era più facile che venisse reputato incapace d'intendere e di volere. I caucasici avevano

il 78% in meno di probabilità di ricevere una diagnosi di disturbo psicotico e, al di là dell'etnia, più erano alti i livelli di educazione e minore era la probabilità che questa diagnosi venisse fatta. Conclusero che livello socioculturale e razza erano elementi centrali per la diagnosi di disturbo psicotico. Gowensmith e McCallum (2019) fanno presente che il livello socioculturale può giocare un ruolo centrale nelle valutazioni forensi anche perché chi è più povero ha una minore disponibilità economica e, quindi, più difficilmente riuscirà a procurarsi un buon difensore. Gli autori sottolineano, inoltre, come spesso i più svantaggiati economicamente siano proprio gli stranieri. Un ulteriore aspetto da considerare è che quando un soggetto che appartiene ad una minoranza viene sottoposto a valutazione può partire dal presupposto che sarà vittima di pregiudizio e, quindi, mettere in atto dei comportamenti che, paradossalmente, favoriscono i *bias* razziali (Hicks, 2004). Altri pregiudizi sono quelli legati alle differenze tra uomo e donna. L'uomo, comunemente, viene considerato più propenso alla violenza rispetto alla donna e, per questo motivo, è più frequente che una donna violenta venga valutata affetta da una grave malattia mentale e quindi non imputabile (Yourstone et al., 2008; Mandarelli et al., 2019). Sygel et al. (2015) ipotizzarono che questo *bias* di genere fosse presente anche in Svezia, ma non ottennero gli stessi risultati. Dal loro studio, infatti, emerse che non c'erano differenze significative legate al genere dell'imputato né tanto meno al genere del valutatore. Gli stessi autori ammisero come limite dell'esperimento la bassa numerosità del campione, ma ipotizzarono che, qualora i risultati fossero stati confermati da altri studi, il motivo di questi esiti sarebbe potuto dipendere dal fatto che, in Svezia, il *National Board of Forensic Medicine* ha il monopolio sulle valutazioni psichiatriche forensi, garantendo una buona qualità procedurale. La Svezia, inoltre, si è sempre dimostrata più evoluta di altre nazioni per quanto riguarda la parità di genere.

Infine, terminato il lavoro dell'esperto sarà suo compito comunicare l'esito del suo processo decisionale. In questa fase, il livello di sicurezza con cui l'esperto esprime il suo parere conduce a reputare il suo lavoro più o meno accurato. Erroneamente, quindi, si giudica il livello di accuratezza di un processo decisionale sulla base del livello di sicurezza dell'esperto. In realtà, però, non c'è alcuna relazione tra questi due aspetti, ma, credere che ci sia può condurre a sua volta chi ascolta ad essere vittima di *bias* (Arkes, 1989).

Varie proposte sono state avanzate al fine di migliorare le valutazioni psichiatriche-forensi e mitigare gli errori sopraelencati. La maggior parte degli autori propone una standardizzazione della procedura in modo da evitare che l'esperto trascuri aspetti importanti del caso (Nicholson & Norwood, 2000; Skellern, 2015; Zapf & Dror, 2017; Gowensmith & McCallum, 2019).

D'altra parte, il problema principale è che l'esperto tende spesso a sottovalutare la possibilità che egli stesso operi un processo decisionale imperfetto ed è, quindi, restio a correggersi o a prestare particolare attenzione al proprio operato. Gowensmith e McCallum definiscono *"bias blind-spot"* le difficoltà che si incontrano nel riconoscere i propri errori rispetto a quelli che commettono gli altri e ipotizzano che gli esperti siano così restii a riconoscere i propri sbagli a differenza di quelli altrui, perché quando si valutano sono più introspettivi e tendono a giustificarsi. Questi limiti rendono l'utilizzo di strumenti dotati di IA una delle strade più interessanti e promettenti al momento al fine di mitigare i problemi sino ad ora elencati, quali ad esempio i *bias* cognitivi, tentando allo stesso tempo di rendere le procedure valutative più standardizzate.

Una soluzione innovativa: l'Intelligenza Artificiale (IA)

Il termine Intelligenza Artificiale indica i sistemi che mostrano un comportamento intelligente analizzando il proprio ambiente e compiendo azioni, con un certo grado di autonomia, per raggiungere specifici obiettivi [Comunicazione del 2018 elaborata dalla Commissione europea, intitolata "*Artificial Intelligence for Europe*"]. Il *Machine Learning* (ML) (o apprendimento automatico) è l'insieme dei metodi di apprendimento che permettono alla macchina dotata di IA di riuscire a comprendere e a risolvere specifiche istanze senza essere stata preventivamente programmata (Mitchell, 1997). La validità di un processo di apprendimento automatico è determinata dall'abilità di generalizzare, ovvero dalla capacità di riuscire a risolvere anche problemi mai esaminati precedentemente sulla base dell'esperienza acquisita (Tortora et al., 2020). Affinché lo strumento dotato di IA diventi accurato è necessaria una fase di addestramento che permetta di acquisire competenze per la risoluzione di specifici problemi.

Sistemi di IA sono stati progettati per riuscire a svolgere in maniera indipendente anche valutazioni psichiatriche-forensi, al fine di ridurre i *bias* cognitivi e rendere le procedure valutative più standardizzate. Particolare interesse ha riscosso l'uso dell'IA al fine di rendere più oggettive le valutazioni relative alla pericolosità dell'imputato (*risk assessment*) (Tortora et al., 2020). Questi sistemi predittivi presuppongono la previa individuazione di una serie di fattori di rischio (o predittori) direttamente coinvolti nel comportamento criminoso. Una volta individuati i fattori che, sulla base della letteratura scientifica più accreditata, aumentano il rischio che l'imputato sia pericoloso, l'IA se ne avvale per risolvere il caso specifico. I dati relativi all'imputato che sembrano giocare un ruolo statisticamente più significativo per la predizione della pericolosità sono: l'età, il sesso, l'origine etnica, il livello di scolarizzazione, la situazione familiare e lavorativa, la posizione sociale, i precedenti penali, le precedenti esperienze carcerarie, i precedenti episodi di violenza agita, le pregresse ospedalizzazioni, il pensiero pro-criminale, alcune variabili contestuali (quali, ad esempio, la mancanza di sostegno familiare e sociale), il consumo di sostanze stupefacenti o alcoliche, la psicopatia (Basile, 2019). Lo strumento dotato di IA, quindi, prima apprende il ruolo che questi

fattori di rischio giocano sulla valutazione della pericolosità e, sulla base del modello appreso, dovrebbe riuscire a generalizzare la conoscenza acquisita e a risolvere in modo autonomo nuovi casi mai visti prima. Con questi strumenti si aspira ad una riduzione di quegli errori procedurali causati da *bias* cognitivi che sono stati elencati nei paragrafi precedenti, a favore di una maggiore standardizzazione procedurale. Queste valutazioni, infatti, si basano su risultati di analisi statistiche.

Negli Stati Uniti d'America, è già molto comune l'utilizzo dell'IA per le valutazioni forensi relative alla pericolosità dell'imputato (Angwin et al., 2016). L'analisi di questi strumenti e la loro applicazione all'interno del processo ha sollevato alcune critiche e ha obbligato i professionisti del settore a riflettere sulle implicazioni che l'utilizzo di questi strumenti può comportare in un contesto forense (Basile, 2019).

In primo luogo, si è dibattuto sul fatto che qualsiasi algoritmo non ha una struttura neutra e lo sviluppatore, in fase di architettura, fa delle scelte che, necessariamente, influenzano il risultato dell'operazione computazionale, ovvero influenzano l'esito del processo decisionale dell'IA. Ad esempio, in fase di progettazione, gli esperti scelgono quali sono gli aspetti che l'IA dovrà prendere in considerazione per risolvere il problema che gli viene posto. Questo fa sì che lo strumento non possa essere considerato privo di *bias*. Inoltre, nonostante algoritmi di predizione del rischio di recidiva del crimine siano percepiti come mezzi per superare i bias umani, essi stessi non sono esenti dal pregiudizio e dal bias istituzionale. Difatti, tali algoritmi vengono generalmente sottoposti ad un training di validazione su dati che possono essi stessi riflettere dei bias (Tortora et al., 2020). Nel momento in cui, poi, l'algoritmo alla base del processo decisionale è protetto da diritti di proprietà intellettuale, ovvero è impedita la divulgazione di informazioni relative al suo metodo di funzionamento, esso sarà sottratto alla possibilità di controllo, verifica e confutazione rendendo di fatto impossibile l'individuazione di eventuali *bias* (Basile, 2019).

La fase della raccolta e della selezione delle variabili da prendere in considerazione per le valutazioni psichiatriche-forensi, inoltre, è particolarmente delicata, soprattutto per valutazioni complesse come quelle relative alla capacità di intendere e di volere dell'autore di reato. In questo caso, al perito viene richiesta una valutazione retrospettiva dello stato di mente dell'imputato al momento del crimine al fine di accertare la presenza di una infermità di mente tale da incidere concretamente sulla capacità di intendere e di volere, escludendola o scemandola grandemente, e a condizione che sussista un nesso eziologico con la specifica condotta criminosa per effetto del quale il fatto reato sia ritenuto causalmente determinato dal disturbo mentale. Il concetto di infermità di mente non è sovrapponibile a quello di malattia mentale e, di conseguenza, la semplice applicazione delle categorie diagnostiche può essere fonte di rischi e fraintendimenti (Gulotta, 2011). Al fine di definire le variabili che l'esperto dovrebbe prendere in considerazione, esistono degli strumenti specifici che guidano e supportano il professionista che è chiamato a rispondere a determinati quesiti posti dal giudice. Rispetto alle valutazioni psichiatriche-forensi relative allo stato di mente dell'imputato al momento del crimine, ad esempio, ricordiamo la Defendant's Insanity Assessment Support Scale (DIASS) (Parmigiani et al., 2019). La DIASS è uno strumento di guida e supporto per lo psichiatra forense chiamato a valutare l'infermità di mente dell'autore di reato al momento del fatto, sviluppato col fine di migliorare l'affidabilità e la coerenza di tali valutazioni. È composto da otto items a cui l'esperto, dopo aver esaminato la documentazione disponibile in merito al caso ed effettuato il colloquio psichiatrico-forense, dovrebbe rispondere per fare una valutazione più accurata dell'imputato in esame (Parmigiani et al., 2019; Parmigiani et al., 2022). Strumenti come questo possono essere sicuramente di ispirazione e di riferimento nella fase di selezione delle variabili che il *tool* dotato di IA dovrebbe prendere in considerazione.

Un secondo aspetto, collegato al precedente, riguarda gli algoritmi *black box*. La espressione *"black box"* sta ad indicare quegli algoritmi che producono un processo decisionale che non è pienamente comprensibile né per le persone coinvolte né per gli esperti di informatica. In pratica, in questi casi è possibile conoscere solo l'esito del processo decisionale mentre alcuni aspetti che hanno determinato quell'esito restano sconosciuti (Hannah-Moffat, 2015).

Nel caso in cui l'algoritmo è *black box* e/o protetto dal segreto professionale viene violato il principio di esplicabilità secondo il quale i processi decisionali dei sistemi di IA devono essere trasparenti, le capacità e lo scopo dei sistemi apertamente comunicati, le decisioni comprensibili a coloro che sono direttamente e indirettamente interessati (Brewka, 1996). Nell'ambito giuridico italiano, il concetto di esplicabilità diventa di particolare rilevanza, considerando che principio fondamentale del processo civile, tributario, penale e amministrativo è il principio del contraddittorio. Esso indica una garanzia di giustizia secondo la quale nessuno può subire gli effetti di una sentenza senza avere avuto la possibilità di essere parte del processo da cui la stessa proviene, ossia senza aver avuto la possibilità di un'effettiva partecipazione alla formazione del provvedimento giurisdizionale. È un principio che implica, quindi, un confronto argomentativo tra posizioni o opinioni diverse, in condizioni di "par condicio". Di conseguenza, l'utilizzo, durante il processo, di un'IA priva di esplicabilità si pone automaticamente in contrasto con il principio del contraddittorio. Sarebbe auspicabile, quindi, che il *software* dotato di IA non fosse protetto dal segreto industriale e fosse in grado, non soltanto di fornire informazioni rispetto al rischio di recidiva del soggetto criminale, ma anche di "spiegare" in che modo è giunto a quella determinata conclusione. Conoscere il processo decisionale dal quale è scaturito quel certo risultato permetterebbe a tutti i protagonisti del processo di poter valutare le ragioni sottostanti a quella decisione e, eventualmente, contrastarle.

Un altro punto oggetto di dibattito è stato quello relativo ai così detti "algoritmi discriminatori". Da alcune analisi, infatti, era emerso che certi strumenti di IA stabilivano se gli imputati erano o meno a rischio di recidiva basandosi principalmente sulla razza e il genere degli imputati stessi (Angwin et al., 2016; Kehl & Kessler, 2017; Washington, 2018). A questo proposito, alcuni esperti hanno evidenziato che il genere e la razza sono variabili che statisticamente favoriscono una valutazione maggiormente accurata del rischio di recidiva e, in quanto tali, non hanno una valenza discriminatoria, perché, nel momento in cui due gruppi vengono trattati diversamente per ragioni scientifiche, questa non può essere definita discriminazione (Corbett-Davies & Goel, 2018). Altri, invece, hanno sostenuto che affermare che aspetti come genere e razza rendono la valutazione più accurata non è una giustificazione abbastanza forte per poterli effettivamente utilizzare nel processo, perché discriminatorio (Kehl & Kessler, 2017). Molti sviluppatori hanno poi deciso di eliminare le variabili genere, livello socioculturale e razza dell'imputato dal processo decisionale degli strumenti dotati di IA usati in ambito forense. D'altra parte, come Corbett-Davies e Goel (2018) hanno fatto notare, non tenere in considerazione differenze tra gruppi che sono statisticamente significative è tutt'altro che equo. Per esempio, è stato dimostrato che le donne sono meno propense a commettere altri crimini in futuro rispetto agli uomini. Come risultato, nel momento in cui l'algoritmo deve valutare gli imputati senza tener conto del genere, sovrastimerà il rischio di recidiva di una donna (Corbett-Davies & Goel, 2018). Gli stessi autori fanno notare che la legge non considera discriminatorie tutte le differenze tra i gruppi, ma solo quelle che conducono ad una disparità ingiustificata. A questo proposito, citano la sentenza Griggs vs. Duke Power Co. con la quale la Corte Suprema aveva chiarito che si viene accusati di avere agito in modo discriminatorio qualora la differenziazione tra gruppi non sia giustificata e non abbia fondamenta razionali. Barabas et al. (2018), invece, all'interno di questa discussione, si pongono ad un altro livello e suggeriscono di non utilizzare l'IA per valutare il rischio di recidiva, ma per individuare le covariate che, in un modello causale, facilitano la comprensione dei fattori sociali, strutturali e psicologici legati ai crimini commessi. Questo tipo di analisi permetterebbe di capire quali sono i fattori sottostanti, per esempio, alle differenze di razza e livello socioculturale che, da un punto di vista statistico, rendono alcuni gruppi di persone più propensi a reiterare il crimine rispetto ad altri. Individuare le covariate e impostare i trattamenti di cura sulla base di esse potrebbe rappresentare un cambiamento sociale e, con il tempo, ridurre le differenze statistiche tra gruppi.

Applicazione dell'IA al campo forense: punti critici e possibili soluzioni

Le valutazioni psichiatriche-forensi sono molto delicate e spesso hanno un ruolo determinante sulle sorti dell'imputato. Assicurarsi che un processo decisionale così complesso sia messo in atto con la dovuta attenzione e cura risulta di fondamentale importanza per garantire la maggiore tutela possibile del singolo, ma anche per evitare una mala gestione dei soldi dello Stato. Per assicurarsi che errori procedurali non si verifichino, la maggior parte degli autori è concorde nel dire che è necessario standardizzare maggiormente le procedure di valutazione (Goldyne, 2007; Gowensmith e McCallum, 2019; Zapf & Dror, 2017), ma gli esperti sono spesso restii a mettere in discussione il proprio operato e sottovalutano quanto esso possa essere determinato da *bias* cognitivi (Gowensmith & McCallum, 2019; Kukucka et al., 2017).

L'IA rappresenta al momento una soluzione innovativa per quanto non priva di problematiche che devono essere affrontate. L'esperto potrebbe avvalersi dello strumento dotato di IA per avere un parere "altro" sul caso in esame. Affinché questo strumento aiuti il professionista a ragionare senza ricorrere ad euristiche, non dovrebbe essere coperto dal segreto industriale e dovrebbe essere progettato in modo da "spiegare" anche il processo decisionale che ha determinato il parere conclusivo (Brewka, 1996; Kehl & Kessler, 2017). Questa caratteristica appare fondamentale per garantire il contraddittorio e per consentire al professionista che si avvale dell'IA di valutare criticamente la decisione emessa e decidere se fare proprio o meno quel ragionamento. In quest'ottica, l'IA rappresenterebbe uno strumento che obbliga l'esperto a operare un ragionamento di tipo controfattuale. Il professionista, dovendo analizzare il processo decisionale della macchina, si troverebbe a confrontarlo con il proprio. Il confronto obbligherebbe di fatto l'esperto a revisionare il proprio ragionamento e a domandarsi se ha approfondito tutti i punti che l'IA ha reputato importanti per la decisione presa o, al contrario, ha dato peso ad altri fattori che avrebbero dovuto essere meno determinati. L'esperto resterebbe comunque libero di dissentire dal parere dell'IA, ma esplicitando i motivi che lo hanno condotto a prendere una decisione differente. Questo passaggio sarebbe fondamentale, perché renderebbe più trasparente non solo il processo decisionale dell'IA ma anche quello dell'esperto stesso. In definitiva, si tratterebbe di un sistema di supporto decisionale dotato di IA.

Ulteriore attenzione merita il momento di selezione delle variabili (fattori di rischio) nella fase di progettazione. Esse sono importanti affinché lo strumento dotato di IA riesca a mettere in atto un buon processo decisionale (Corbett-Davies & Goel, 2018). A nostro avviso, l'IA non dovrebbe essere protetta da diritti di proprietà intellettuale in modo da consentire il controllo, la verifica e la confutazione dell'algoritmo. Inoltre, riteniamo che in fase di progettazione bisognerebbe tenere conto anche delle variabili che fanno riferimento a genere, razza e liv-

ello socioculturale qualora dati statistici ne dimostrino la rilevanza ai fini della decisione. La figura dell'esperto, però, rimarrebbe fondamentale. L'IA sarebbe in grado di produrre un output basandosi "soltanto" sulle variabili che solitamente sono importanti, perdendo, però, tutti quegli aspetti peculiari del caso in esame che spetterà al valutatore identificare. Lo strumento potrebbe snellire il lavoro del perito e del consulente, assicurare una maggiore standardizzazione della procedura, obbligare l'esperto a mettere in discussione il proprio giudizio.

In parallelo, strumenti di IA dovrebbero essere sviluppati al fine di agire ad un livello più profondo, favorendo l'individuazione delle covariate del comportamento criminale e aiutando a programmare l'intervento migliore per i soggetti maggiormente a rischio di commettere atti criminali (Barabas et al., 2018). Significherebbe agire su quei fattori che rendono alcune categorie più propense di altre a commettere crimini, ovvero intervenire su quelle variabili che determinano le sistematiche differenze di razza, genere e livello socioculturale e, a lungo termine, non renderle più discriminanti per le decisioni dell'IA.

Inoltre, un altro aspetto che, a nostro avviso, varrebbe la pena considerare sono le differenze culturali. Ogni struttura giuridica, infatti, è imbevuta della cultura del suo popolo e, per questa ragione, l'IA dovrebbe essere costruita tenendo conto della popolazione specifica sulla quale andrà poi applicata e delle leggi che caratterizzano quella nazione (Kehl & Kessler, 2017).

Applicazione dell'IA agli altri campi della medicina: riflessioni e proposte

I tentativi di applicazione dell'IA alla psichiatria forense sono ancora pochi e poco soddisfacenti. Sembra che sia ancora molta la strada da fare affinché questi strumenti possano garantire un processo decisionale realmente privo di *bias*. Dall'altro lato, osservando il potenziamento degli strumenti dotati di IA nell'area della medicina in generale possiamo ben sperare che lo stesso avvenga anche nel campo della psichiatria forense. Le linee guida redatte dalla Commissione Europea, infatti, incentivano l'utilizzo dell'IA in ambito medico e lo considerano uno dei settori in cui si possono ottenere risultati più interessanti e rivoluzionari (AI, 2019). Studiare gli strumenti di IA che vengono usati in questo contesto ci può aiutare a proporre nuove soluzioni volte a migliorare l'applicazione di questi strumenti al campo forense. In campo medico, ad esempio, esistono già casi di sistemi di supporto decisionale dotati di IA che permettono al medico di operare un processo decisionale più trasparente e di assicurarsi di aver preso in considerazione tutti gli aspetti importanti ai fini della decisione (Cruz Rivera et al., 2020; Keller et al., 2020). Sono stati anche sviluppati strumenti che consentono sperimentazioni cliniche eseguite mediante simulazioni al computer (i cosiddetti *in silico clinical trials*, cfr. ad es. Avicenna, 2016; European Medicines Agency [EMA], 2018; Food and Drug Administration [FDA], 2018; Maggioli et al., 2020; Pappalardo et al., 2019) di modelli matematici della fisiologia umana di interesse (cfr. ad es., Hester et al, 2011), della cinetica e della dinamica dei farmaci (cfr. ad es., Lippert et al., 2019) su una popolazione di *pazienti virtuali* (Sinisi et al., 2020a). Tali popolazioni possono essere usate per generare gemelli virtuali (*digital twin*, cfr. Sinisi et al., 2020b) dei pazienti umani da trattare. Questo abilita l'impiego di tecniche di IA per supportare il clinico nella scelta di trattamenti specificatamente *individualizzati* su (ed *ottimizzati* per) il singolo paziente, ovvero che, in base alle peculiarità di quest'ultimo, massimizzino l'efficacia del trattamento, mantengano basso il rischio di effetti avversi e garantiscano il rispetto delle linee guida. I trattamenti così calcolati non sono frutto di *bias* cognitivi (come spesso avviene con i processi decisionali umani) né di statistiche apprese considerando *insiemi* di pazienti (come avviene quando si utilizzano approcci di IA basati esclusivamente su ML). Questi trattamenti vengono testati sul *singolo* paziente in esame in modo oggettivo, mediante simulazioni sul suo gemello virtuale (compatibilmente con il grado di confidenza raggiunto su quest'ultimo), anche tenendo conto della possibile incertezza nelle misure cliniche effettuate e della probabilità di eventi esogeni incontrollabili, come avviene tipicamente nelle scienze fisiche e nell'ingegneria (cfr., ad es., Mancini et al, 2018; Mancini et al., 2021). Sebbene ancora allo stato iniziale, tali approcci alla medicina di precisione basati su modelli fisiologici individualizzati sono già stati applicati con successo. Ad esempio, questo è avvenuto nel campo dell'endocrinologia riproduttiva (Fischer et al, 2021), un'area che vede la presenza di molti fattori difficili da tenere simultaneamente in debita considerazione (Hengartner et al., 2017; Leeners et al., 2017; Leeners et al., 2019; Leeners et al., 2021). Chiaramente l'applicazione di tali metodi al supporto decisionale in ambito forense non è affatto immediata; tuttavia, essi possono essere d'ispirazione e rappresentare un buon punto di partenza. Potremmo aspirare allo sviluppo di modelli comportamentali qualitativi e di alto livello che non si occupino di generare previsioni ma che indirizzino l'esperto forense nella sua pratica aiutandolo a non prendere quelle decisioni che si rivelino (anche indirettamente) in contrasto con conoscenze consolidate.

In conclusione, individuati i problemi che hanno contraddistinto i tentativi di applicazione dell'IA al campo forense fino ad ora, sarebbe opportuno iniziare ad esplorare nuove strade per implementare sistemi di supporto decisionale che siano più esplicabili e trasparenti possibile. Osservare l'evoluzione che l'IA sta avendo negli altri campi della medicina potrebbe rappresentare un buon punto di partenza per perfezionarne l'applicazione anche in un contesto forense.

Riferimenti bibliografici

AI, H. (2019). High-level expert group on artificial intelligence. *Ethics guidelines for trustworthy AI*. DOI: 10.2759/346720

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks. *ProPublica*. Retrieved December 23, 2016.

Arkes, H. R. (1989). Principles in judgment/decision making research pertinent to legal proceedings. *Behavioral Sciences & the Law*, *7*(4), 429-456. https://doi.org/10.1002-/bsl.2370070403

Avicenna (2016). *In silico clinical trials*. Ottenuto da avicenna-isct.org

Barabas, C., Virza, M., Dinakar, K., Ito, J., & Zittrain, J. (2018, January). Interventions over predictions: Reframing the ethical debate for actuarial risk assessment. In *Conference on Fairness, Accountability and Transparency* (pp. 62-76). PMLR. https://doi.org/10.48550/arXiv.1712.08238

Basile, F. (2019). *Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine*. DOI: 10.1007/s13347-019-00345-yp)

Blow, F. C., Zeber, J. E., McCarthy, J. F., Valenstein, M., Gillon, L., & Bingham, C. R. (2004). Ethnicity and diagnostic patterns in veterans with psychoses. *Social Psychiatry and Psychiatric Epidemiology*, 39, 841–851. DOI:10.1007/s00127-004-0824-7

Brewka, G. (1996). Artificial intelligence—a modern approach by Stuart Russell and Peter Norvig, Prentice Hall. Series in Artificial Intelligence, Englewood Cliffs, NJ. *The Knowledge Engineering Review*, 11(1), 78-79) DOI: https://doi.org/-10.1017/S0269888900007724

Cooper, G. S., & Meterko, V. (2019). Cognitive bias research in forensic science: a systematic review. *Forensic science international*, *297*, 35-46. DOI: 10.1016/j.forsciint.2019.01.016

Corbett-Corbett-Davies, S., & Goel, S. (2018). The measure and mismeasure of fairness: A critical review of fair machine learning. *arXiv preprint arXiv:1808.00023*. https://doi.org/-10.48550/arXiv.1808.00023

Cruz Rivera, S., Liu, X., Chan, A. W., Denniston, A. K., & Calvert, M. J. (2020). Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *Nature medicine*, *26*(9), 1351-1363. DOI: 10.1016/S2589-7500(20)30219-3

Curley, L. J., Munro, J., & Dror, I. E. (2022). Cognitive and human factors in legal layperson decision making: Sources of bias in juror decision making. *Medicine, Science and the Law*, 00258024221080655. DOI: 10.1177/0025802-4221080655

Dror, I. E. (2020). Cognitive and human factors in expert decision making: six fallacies and the eight sources of bias. *Analytical Chemistry*, *92*(12), 7998-8004. DOI: 10.1021/acs.analchem.0c00704

EMA. (2018). Guideline on the reporting of physiologically based pharmacokinetic (PBPK) modelling and simulation.

Everson, M. D., & Sandoval, J. M. (2011). Forensic child sexual abuse evaluations: Assessing subjectivity and bias in professional judgements. *Child Abuse & Neglect*, *35*(4), 287-298. https://doi.org/10.1016/j.chiabu.2011.01.001

FDA, U. (2018). Physiologically based pharmacokinetic analyses: format and content, guidance for industry.

Fischer, S., Ehrig, R., Schäfer, S., Tronci, E., Mancini, T., Egli, M., ... & Röblitz, S. (2021). Mathematical modeling and simulation provides evidence for new strategies of ovarian stimulation. *Frontiers in endocrinology*, *12*, 117. DOI: 10.3389/fendo.2021.613048

Goldyne, A. J. (2007). Minimizing the influence of unconscious bias in evaluations: A practical guide. *Journal-American Academy Of Psychiatry And The Law*, *35*(1), 60.

Gowensmith, W. N., & McCallum, K. E. (2019). Mirror, mirror on the wall, who's the least biased of them all? Dangers and potential solutions regarding bias in forensic psychological evaluations. *South African journal of psychology*, *49*(2), 165-176. https://doi.org/10.1177/0081246-319835117

Gulotta, G. (2011). *Compendio di psicologia giuridico-forense, criminale e investigativa* (Vol. 53). Giuffrè Editore.

Hannah-Moffat, K. (2015). The uncertainties of risk assessment: Partiality, transparency, and just decisions. *Federal Sentencing Reporter*, *27*(4), 244-247. https://doi.org/10.1525/fsr.2-015.27.4.244

Haselton, M. G., Nettle, D., & Murray, D. R. (2015). The evolution of cognitive bias. *The handbook of evolutionary psychology*, 1-20. https://doi.org/10.1002/9781119125-563.evpsych241

Hengartner, M. P., Kruger, T. H., Geraedts, K., Tronci, E., Mancini, T., Ille, F., ... & Leeners, B. (2017). Negative affect is unrelated to fluctuations in hormone levels across the menstrual cycle: Evidence from a multisite observational study across two successive cycles. *Journal of psychosomatic research*, *99*, 21-27. DOI: 10.1016/j.jpsychores.2017.05.018

Hester, R., Brown, A., Husband, L., Iliescu, R., Pruett, W. A., Summers, R. L., & Coleman, T. (2011). HumMod: a modeling environment for the simulation of integrative human physiology. *Frontiers in physiology*, *2*, 12. DOI: 10.3-389/fphys.2011.00012

Hicks, J. W. (2004). Ethnicity, race, and forensic psychiatry: are we color-blind?. *Journal of the American Academy of Psychiatry and the Law Online*, *32*(1), 21-33.

Kehl, D. L., & Kessler, S. A. (2017). *Algorithms in the criminal justice system: Assessing the use of risk assessments in sentencing*.

Keller, N., Jenny, M. A., Spies, C. A., & Herzog, S. M. (2020, November). Augmenting Decision Competence in Healthcare Using AI-based Cognitive Models. In *2020 IEEE International Conference on Healthcare Informatics (ICHI)* (pp. 1-4). IEEE. DOI: 10.1109/ICHI48887.2020.9374376

Kukucka, J., Kassin, S. M., Zapf, P. A., & Dror, I. E. (2017). Cognitive bias and blindness: a global survey of forensic science examiners. *Journal of Applied Research in Memory and Cognition*, *6*(4), 452-459. https://doi.org/10.1016/j.jarmac.2017.09.001

Leeners, B., Kruger, T. H., Geraedts, K., Tronci, E., Mancini, T., Ille, F., ... & Hengartner, M. P. (2017). Lack of associations between female hormone levels and visuospatial working memory, divided attention and cognitive bias across two consecutive menstrual cycles. *Frontiers in behavioral neuroscience*, 120.DOI: 10.3389/fnbeh.2017.00120.

Leeners, B., Krüger, T. H., Geraedts, K., Tronci, E., Mancini, T., Egli, M., ... & Ille, F. (2019). Associations between natural physiological and supraphysiological estradiol levels and stress perception. *Frontiers in Psychology*, *10*, 1296. DOI: 10.3389/fpsyg.2019.01296.

Leeners, B., Krüger, T., Geraedts, K., Tronci, E., Mancini, T., Ille, F., ... & Hengartner, M. P. (2021). Cognitive function in association with high estradiol levels resulting from fertility treatment. *Hormones and behavior*, *130*, 104951. DOI: 10.1016/j.yhbeh.2021.104951.

Lippert, J., Burghaus, R., Edginton, A., Frechen, S., Karlsson, M., Kovar, A., ... & Teutonico, D. (2019). Open systems pharmacology community—an open access, open source, open science approach to modeling and simulation in pharmaceutical sciences. *CPT: pharmacometrics & systems pharmacology*, *8*(12), 878. DOI: 10.1002/psp4.12473

Maggioli, F., Mancini, T., & Tronci, E. (2020). SBML2Modelica: integrating biochemical models within open-standard simulation ecosystems. *Bioinformatics*, *36*(7), 2165-2172. DOI: 10.1093/bioinformatics/btz860

Mancini, T., Mari, F., Melatti, I., Salvo, I., Tronci, E., Gruber, J. K., ... & Elmegaard, L. (2018, October). Parallel statistical model checking for safety verification in smart grids. In *2018 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)* (pp. 1-6). IEEE.. DOI: 10.1109/SmartGridComm.2018.8587416

Mancini, T., Melatti, I., & Tronci, E. (2021). Any-horizon uniform random sampling and enumeration of constrained scenarios for simulation-based formal verification. *IEEE Transactions on Software Engineering*. DOI: 10.1109/TSE.2021.3109842

Mandarelli G., Carabellese F., Felthous A.R., et al. (2019). The factors associated with forensic psychiatrists' decisions in criminal responsibility and social dangerousness evaluations. *Int J Law Psychiatry*, 66, 101503. Doi:10.1016/j.ijlp.2019.101503

Mikton, C., & Grounds, A. (2007). Cross-cultural clinical judgment bias in personality disorder diagnosis by forensic psychiatrists in the UK: A case-vignette study. *Journal of personality disorders*, *21*(4), 400-417. https://doi.org/10.1521/pedi.2007.21.4.400

Mitchell, T. M. (1997). *Machine learning*. Burr Ridge, IL: McGraw Hill

Neal, T., & Grisso, T. (2014). The cognitive underpinnings of bias in forensic mental health evaluations. *Psychology, Public Policy, and Law*, 202, 200–211. DOI:10.1037/a0035824

Neighbors, H. W., Trierweiler, S. J., Munday, C., Thompson, E. E., Jackson, J. S., Binion, V. J., et al. (1999). Psychiatric diagnosis of African Americans: Diagnostic divergence in clinicianstructured and semistructured interviewing conditions. *Journal of the National Medical Association*, 91, 601–612.

Nicholson, R. A., & Norwood, S. (2000). The quality of forensic psychological assessments, reports, and testimony: Acknowledging the gap between promise and practice. *Law and human Behavior*, *24*(1), 9-44. DOI: 10.1023/A:1005422702678

Pappalardo, F., Russo, G., Tshinanu, F. M., & Viceconti, M. (2019). In silico clinical trials: concepts and early adoptions. *Briefings in bioinformatics*, *20*(5), 1699-1708. DOI: 10.1093/bib/bby043.

Parmigiani G., Mandarelli G., Meynen G., Carabellese F., Ferracuti S. (2019). Translating clinical findings to the legal norm: the Defendant's Insanity Assessment Support Scale (DIASS). *Transl Psychiatry*, 9(1), 278. Published 2019 Nov 7. doi:10.1038/s41398-019-0628-x

Parmigiani G., Mandarelli G., Roma P., Ferracuti S. (2022). Validation of a new instrument to guide and support insanity evaluations: the defendant's insanity assessment support scale (DIASS). *Transl Psychiatry*, 12(1), 115. Published 2022 Mar 22. doi:10.1038/s41398-022-01871-8.

Perry, B. L., Neltner, M., & Allen, T. (2013). A paradox of bias: Racial differences in forensic psychiatric diagnosis and determinations of criminal responsibility. *Race and social problems*, *5*(4), 239-249. DOI: 10.1007/s12552-013-9100-3

Sattar, S. P., Pinals, D. A., & Gutheil, T. (2002). Countering countertransference: a forensic trainee's dilemma. *Journal of the American Academy of Psychiatry and the Law Online*, *30*(1), 65-69.

Sinisi, S., Alimguzhin, V., Mancini, T., Tronci, E., & Leeners, B. (2020a). Complete populations of virtual patients for in silico clinical trials. *Bioinformatics*, *36*(22-23), 5465-5472. https://doi.org/10.1093/bioinformatics/btaa1026

Sinisi, S., Alimguzhin, V., Mancini, T., Tronci, E., Mari, F., & Leeners, B. (2020b). Optimal personalised treatment computation through in silico clinical trials on patient digital twins. *Fundamenta Informaticae*, *174*(3-4), 283-310. DOI: 10.3233/FI-2020-1943

Skellern, C. (2015). Minimising bias in the forensic evaluation of suspicious paediatric injury. *Journal of forensic and legal medicine*, *34*, 11-16. https://doi.org/10.1016/j.jflm.2015.05.002

Stracciari, A., Caucasici, A., & Sartori, G. (2010). *Neuropsicologia forense*. Il mulino.

Strakowski, S. M., Keck, P. E, Jr, Arnold, L. M., Collins, J., Wilson, R. M., Fleck, D. E., et al. (2003). Ethnicity and diagnosis in patients with affective disorders. Journal of Clinical Psychiatry, 64(7), 747–754 DOI: 10.4088/JCP.v64n0702

Sygel, K., Sturup, J., Fors, U., Edberg, H., Gavazzeni, J., Howner, K., ... & Kristiansson, M. (2017). The effect of gender on the outcome of forensic psychiatric assessment in Sweden: A case vignette study. *Criminal behaviour and mental health*, *27*(2), 124-135. DOI: 10.1002/cbm.1987

Tortora L, Meynen G, Bijlsma J, Tronci E, Ferracuti S. Neuroprediction and A.I. in Forensic Psychiatry and Criminal Justice: A Neurolaw Perspective. Front Psychol. 2020 Mar 17;11:220. DOI: 10.3389/fpsyg.2020.00220. PMID: 32256422; PMCID: PMC7090235.

Washington, A. L. (2018). How to argue with an algorithm: Lessons from the COMPAS-ProPublica debate. *Colo. Tech. LJ*, *17*, 131.

Yourstone, Lindholm, Grann M e Svenson, 2008) Evidence of gender bias in legal insanity evaluations: a case vignette study of clinicians, judges and students. Nordic Journal of Psychiatry 62: 273–278. DOI:10.1080/08039480801963135

Zapf, P. A., & Dror, I. E. (2017). Understanding and mitigating bias in forensic evaluation: lessons from forensic science. *International Journal of Forensic Mental Health*, *16*(3), 227-238. https://doi.org/10.1080/14999013.2017.1317302