ISSN 1826-3534

federalismi.it
RIVISTA DI DIRITTO PUBBLICO ITALIANO, COMPARATO, EUROPEO

FOCUS LAVORO, PERSONA, TECNOLOGIA
18 DICEMBRE 2024

IA e obblighi datoriali di tutela del lavoratore: necessità e declinazioni dell'approccio *risk-based*

di Marco Peruzzi
Professore associato di Diritto del lavoro
Università degli Studi di Verona

IA e obblighi datoriali di tutela del lavoratore: necessità e declinazioni dell'approccio *risk-based*[*]

di Marco Peruzzi
Professore associato di Diritto del lavoro
Università degli Studi di Verona

Abstract [It]: L'analisi evidenzia l'importanza dell'approccio basato sul rischio per l'effettività dei diritti in un contesto di utilizzo dell'IA, esaminando il contributo e i limiti che offrono al riguardo le diverse fonti coinvolte nella costruzione del sistema integrato delle tutele. Proprio la centralità dell'approccio nel contesto in esame, segnalata ad esempio dai rischi di c.d. *proxy discrimination*, invita, d'altra parte, a riflettere sulla rilevanza di una sua implementazione anche al di fuori del campo di applicazione delle diposizioni che lo prevedono. Specifica attenzione è infine dedicata alla tecnica di giuridificazione del rischio (inaccettabile, alto e non-alto) utilizzata dall'*AI Act* e alle problematiche interpretative poste dalle tecniche di classificazione.

Title: AI and employer's obligations to protect the worker: needs and declinations of the risk-based approach
Abstract [En]: The analysis highlights the importance of the risk-based approach for the effectiveness of rights in a context of AI use, examining the contribution and limits offered in this regard by the various sources involved in the construction of the integrated system of protections. The centrality of the approach in the AI context, shown for example by the risks of so-called proxy discrimination, invites reflection on the relevance of its implementation also outside the scope of the provisions that envisage it. Finally, specific attention is devoted to the legal regulation of risks (unacceptable, high and non-high) in the AI Act and to the interpretative problems raised by the classification techniques.

Parole chiave: Intelligenza artificiale; approccio basato sul rischio; diritti fondamentali; protezione dei dati personali; proxy discrimination; riconoscimento delle emozioni
Keywords: Artificial intelligence; risk-based approach; fundamental rights; personal data protection; proxy discrimination; emotion recognition

Sommario: 1. IA e tutela del lavoratore: l'importanza dell'approccio *risk-based*. **1.1.** Alla ricerca di un obbligo trasversale di valutazione di impatto in un sistema di fonti a geometria variabile. **1.2.** La necessità di integrare l'approccio *risk-based* per garantire l'effettività dei diritti in un contesto di utilizzo dell'IA: l'esempio della *proxy discrimination*. La prospettiva volontaristica. **2.** Le tecniche di giuridificazione del rischio nell'*AI Act* e i problemi derivanti dall'incerta delimitazione delle fattispecie normative.

1. IA e tutela del lavoratore: l'importanza dell'approccio *risk-based*

Ragionare di obblighi datoriali di tutela del lavoratore rispetto all'uso di sistemi di IA significa necessariamente ragionare di sistema integrato di fonti.

È un'operazione a cui invita lo stesso Regolamento (Ue) 2024/1689, c.d. *AI Act*, nel momento in cui si qualifica come livello normativo non solo minimo (v. art. 2, par. 11), ma anche complementare, ossia

[*] Articolo sottoposto a referaggio.

dichiaratamente volto ad affiancare le altre fonti senza pregiudicarne l'applicazione[1], nonché a svolgere un ruolo funzionale e abilitante nei confronti dei rimedi e diritti esistenti.

Tale complementarità è espressamente ribadita proprio con riguardo alla figura del *deployer* (qualificazione assunta, come noto, dal datore che faccia uso di un sistema di IA sotto la propria autorità), laddove si chiarisce che gli obblighi stabiliti a suo carico dal Regolamento, in particolare quello di dare applicazione alle istruzioni d'uso e misure di sorveglianza umana indicate dal fornitore, non pregiudicano gli altri obblighi previsti dalle fonti Ue o interne (art. 26, par. 3).

A guidare questo processo di integrazione sistematica sono le scelte di fondo che connotano la risposta regolativa alle sfide della IA, a livello Ue e non solo[2]. Segnatamente, le scelte di contenuto, ossia l'individuazione, quali garanzie imprescindibili di una prospettiva antropocentrica, della trasparenza e supervisione umana, stabiliscono i meridiani su cui vanno cercati i punti di raccordo e saldatura tra i plessi regolativi coinvolti. La scelta di metodo, ossia l'utilizzo dell'approccio basato sul rischio, fornisce la tecnica normativa di costruzione della struttura architettonica.

La scelta dell'approccio *risk-based*, promossa e condivisa anche dalle parti sociali di livello Ue[3], si giustifica a fronte delle caratteristiche del fenomeno da regolare.

Si consideri, anzitutto, che tale tecnica normativa, nel tradurre «una prospettiva procedurale del diritto», rappresenta una «modalità di reazione del diritto ad un elevato grado di differenziazione» e di complessità del reale, tale da impedire una predeterminazione rigida e standard dei contenuti delle misure di tutela[4]. Si tratta, pertanto, di un modello regolatorio di risposta all'inadeguatezza che può presentare la regolazione diretta per misure di dettaglio "nominate" di fronte alla dinamicità evolutiva dei fattori di pericolo e alla modularità degli elementi che li caratterizzano.

---

[1] Il Cons. n. 9 chiarisce che «le norme dovrebbero lasciare impregiudicato il diritto dell'Unione vigente ... rispetto al quale il presente regolamento è complementare. […] il presente regolamento mira a rafforzare l'efficacia di tali diritti e mezzi di ricorso esistenti […]». Lo stesso art. 5, par. 8, chiarisce che la lista di proibizioni di pratiche a rischio inaccettabile ivi prevista lascia impregiudicati i divieti che si applicano qualora una pratica di IA violi altre disposizioni di diritto Ue.

[2] Cfr. la Convenzione del Consiglio d'Europa (di cui fanno parte tutti i 27 Stati membri dell'Ue), completata a marzo 2024 e firmata il 5 settembre 2024 (tra gli altri) dall'Ue, gli Stati Uniti, il Regno Unito e Israele, e il Processo di Hiroshima. Quest'ultimo, condotto dai leader del G7, ha portato, prima, il 30 ottobre 2023, all'adozione di principi-guida internazionali e di un codice di condotta internazionale per le organizzazioni che sviluppano AI avanzata, poi, il 13 settembre 2024, alla sottoscrizione di un piano d'azione, in cui si valorizzano, tra i vari profili, da un lato, il principio di *accountability*, in base al quale «la responsabilità per le decisioni concernenti i rapporti di lavoro, anche qualora prese da sistemi di IA, dovrebbe rimanere in capo ai datori di lavoro», dall'altro, l'importanza del dialogo sociale e della contrattazione collettiva a tutti i livelli.

[3] Cfr. l'accordo quadro sulla digitalizzazione, concluso il 22 giugno 2020 dall'Etuc, per parte sindacale, Businesseurope, Ceep e SMEunited, per parte datoriale, nel contesto del sesto programma di lavoro per il 2019-2021, nonché i testi del dialogo sociale europeo settoriale che sono seguiti, in particolare: la dichiarazione congiunta per il settore delle telecomunicazioni del novembre 2020, le conclusioni congiunte per il settore MET (Metal, Engineering and Technology-based) industries del 22 febbraio 2023 e la dichiarazione congiunta per il settore bancario del 14 maggio 2024.

[4] P. LOI, *La sicurezza. Diritto e fondamento dei diritti nel rapporto di lavoro*, Giappichelli, 2000, pp. 47-48.

Se, come insegna il diritto della sicurezza sul lavoro, la tecnologia è sempre stata una determinante chiave di questa complessità, nel caso dell'IA, soprattutto del *machine learning*, la problematica si intensifica e acuisce, perché le peculiarità adattive e auto-evolutive della fonte di potenziale pericolo, nonché la variabilità della sua possibile autonomia d'azione fanno dipendere l'effettività delle tutele da un necessario processo iterativo di valutazione e gestione dei rischi. Sempre le caratteristiche dell'IA - ossia la dipendenza dai dati per l'applicazione delle tecniche di inferenza, l'opacità del processo di ottenimento dell'*output* e la capacità di quest'ultimo di influenzare gli ambienti fisici e virtuali[5] - impongono che il modello di controllo del rischio si muova sul duplice piano della *governance* dei dati e della verifica periodica di impatto dei processi automatizzati. Ciò è tanto più vero all'aumentare della profondità delle reti neurali e dell'apprendimento automatico, considerato che in tal caso l'investigazione può realizzarsi solo attraverso un'osservazione dall'esterno dell'algoritmo, nel suo funzionamento e nei risultati che consegna[6].

La scelta di metodo, nella costruzione del sistema integrato delle fonti, si confronta, d'altra parte, nella prospettiva lavoristica con la necessità di «delimitazione di zone di indisponibilità, costituite dai diritti sociali fondamentali»[7]. In altre parole, se il paradigma procedurale dell'approccio *risk-based* si presenta funzionale a garantire l'effettività dei diritti nel descritto scenario di complessità, la definizione di tali diritti deve essere presidiata a livello di fonti legislative e costituzionali, in modo che solo queste «rappresentino l'adeguato spazio giuridico in cui effettuare il bilanciamento degli interessi coinvolti»[8]. È in tale spazio che deve essere cioè individuato il grado di tutela dei diritti, venendo affidata all'attore privato, titolare dell'obbligo di valutazione e gestione, solo la scelta delle misure tecniche e organizzative, ossia dei metodi e degli strumenti, che meglio lo garantiscono nello specifico contesto d'incidenza della fonte di pericolo.

---

[5] Cfr. la definizione di «sistema di IA» prevista dall'art. 3, n. 1, *AI Act*. Per una riflessione sulla porosità del confine tracciato da tale definizione, si consenta di rinviare a M. PERUZZI, *Intelligenza artificiale e lavoro: l'impatto dell'AI Act nella ricostruzione del sistema regolativo Ue di tutela*, in M. BIASI (a cura di), *Diritto del lavoro e intelligenza artificiale*, Giuffré, Milano, 2024, pp. 115 ss. Si ricorda come la definizione includa gli approcci basati sulla logica (accanto a quelli di *machine learning*) e prevede in termini solo eventuali l'adattabilità del sistema. Si può segnalare, a tal proposito, come, a livello interno, il d.m 21 novembre 2024 n. 174, nel disciplinare l'utilizzo di strumenti di IA da parte della piattaforma SIISL (Sistema Informativo per l'Inclusione Sociale e Lavorativa) per l'abbinamento ottimale delle domande e offerte di lavoro inserite, escluda espressamente che i dati forniti dall'utente e dalle aziende siano utilizzati per perfezionare il modello, ciò per evitare che dati di bassa qualità possano influenzare il meccanismo di funzionamento dell'algoritmo (art. 14, comma 4, lett. c).

[6] G.F. ITALIANO-S. CIVITARESE MATTEUCCI-A. PERRUCCI, *L'intelligenza artificiale: dalla ricerca scientifica alle sue applicazioni. Una introduzione di contesto*, in A. PAJNO-F. DONATI-A. PERRUCCI (a cura di), *Intelligenza artificiale e diritto: una rivoluzione? Diritti fondamentali, dati personali e regolazione*, vol. I, il Mulino, Bologna, 2022, pp. 43 ss., spec. p. 54.

[7] P. LOI, *La sicurezza*…, cit., p. 48.

[8] P. LOI, *La sicurezza*…, cit., pp. 68-69.

Il modello previsto dalla direttiva 89/391/CEE, concernente la salute e sicurezza durante il lavoro, rappresenta l'esempio di matrice lavoristica della convivenza di questi due paradigmi[9]. È un esempio che salda il grado di tutela dei diritti a un principio di massima sicurezza tecnologica[10], sganciandolo «da considerazioni di carattere puramente economico» (cons. n. 13), e rispetta la necessaria proiezione collettiva delle tutele richiesta dalla materia, prevedendo il coinvolgimento di rappresentanze specializzate nel processo di valutazione e gestione del rischio. Significativamente, all'interno della direttiva 89/391 la combinazione dei due paradigmi incide sulla stessa declinazione dell'approccio basato sul rischio, essendo questo informato dalla prospettiva della c.d. prevenzione primaria, ossia di un modello di tutela dei diritti fondamentali che incorpora «un bilanciamento già intervenuto per opera del legislatore», prevedendo «la tensione alla radicale eliminazione del rischio» e «la copertura surrogatoria in senso limitativo o riduttivo della dose di rischio ineliminabile»[11]. In tal senso, il paradigma procedurale è segnato da una precisa sequenza di obiettivi: prima l'eliminazione del rischio; in subordine, la riduzione al minimo dei rischi ineliminabili. Il datore deve vagliare le scelte organizzative e le possibilità di insorgenza del rischio che vi si ricollegano, al fine di evitare i rischi con adeguate misure di prevenzione; quindi, è tenuto a valutare, gestire e ridurre al minimo i rischi che non si possono evitare, attraverso adeguate misure di protezione [12].

[9] Cfr. P. LOI, *La sicurezza*…, cit. A livello di ordinamento interno, rispetto al «valore, anche di orientamento assiologico, della Carta costituzionale[,] l'unica in grado di riempire di contenuto la formula impiegata dall'art. 2087 c.c. per individuare i beni tutelati attraverso l'imposizione dell'obbligo di sicurezza» cfr. P. ALBI, *Il contenuto dell'obbligo di sicurezza*, in *VTDL*, 2023, n. 4, p. 881.

[10] Si rinvia ad altra sede l'analisi del dibattito, mai sopito, sul livello di massima sicurezza richiesto, se tecnologicamente possibile o fattibile. Cfr. *ex multis* M. GIOVANNONE, *Responsabilità datoriale e prospettive regolative della sicurezza sul lavoro*, Giappichelli, Torino 2024; L. ANGELINI, La sicurezza del lavoro nell'ordinamento europeo, WP Olympus, 2013, n. 29; O BONARDI, *La Corte di Giustizia e l'obbligo di sicurezza del datore di lavoro: il criterio del* reasonably practicable *è assolto per insufficienza di prove*, in *RIDL*, 2008, II, p. 13 ss.; G. NATULLO, *La «massima sicurezza tecnologica»*, in *Diritto e pratica del lavoro*, 1997, p. 815. Si consenta solo di evidenziare come all'interno della direttiva 89/391 si richieda al datore di adeguare il lavoro all'uomo (art. 6, par. 2, lett. d), informandosi «circa i progressi tecnici e le conoscenze scientifiche» (cons. n. 14), e di «tener conto del grado di evoluzione della tecnica» (art. 6, par. 2, lett. e). La stessa Corte di giustizia ha dato delle disposizioni una interpretazione rigorosa, ricordando, ad esempio, che i rischi professionali che devono essere oggetto di una valutazione da parte dei datori di lavoro […] si evolvono costantemente in funzione […] delle ricerche scientifiche in materia» (Corte giust., 15 novembre 2001, *Commissione v. Repubblica italiana*, C 49/00, punto 13). Come è stato spiegato da C. SMURAGLIA, ciò impone al datore «di tenersi aggiornato e di tener conto di quanto risulta da acquisizioni tecniche e scientifiche che abbiano un margine sufficiente di solidità, di sperimentazione e di effettiva possibilità di conoscenza al di là del mondo scientifico, *strictu sensu* considerato» (*Sicurezza del lavoro e obblighi comunitari. I ritardi dell'Italia nell'adempimento e le vie per uscirne*, in *Riv. it. dir. lav.*, 2002, I, pp. 190-191). Rispetto all'impatto, a livello interno, della sentenza della Corte Cost. 25.7.1996 n. 312, se E. GRAGNOLI ritiene «sarebbe singolare se il livello di tutela dipendesse dalla prassi, che l'art. 2087 c.c. vuole migliorare» (*art. 2087*, in F. CARINCI, E. GRAGNOLI (a cura di), *Codice commentato della sicurezza sul lavoro*, Utet, Torino, 2010, p. 45), A. VALLEBONA valorizza la sentenza nell'ottica di una riduzione dei margini di operatività del principio entro i confini del "livello di sicurezza generalmente praticato nel settore" (*Breviario di diritto del lavoro*, Giappichelli, Torino, 2015, p. 283).

[11] S. BUOSO, *Principio di prevenzione e sicurezza sul lavoro*, Giappichelli, Torino, 2020, pp. 48-49.

[12] P. PASCUCCI, *La tutela della salute e della sicurezza sul lavoro. Il Titolo I del d.lgs. n. 81/2008 dopo il Jobs Act*, Aras edizioni, Fano, 2017; C. LAZZARI-P. PASCUCCI, *Sistemi di IA, salute e sicurezza sul lavoro: una sfida al modello di prevenzione aziendale, fra responsabilità e opportunità*, in *RGL*, 2024, n. 4, in corso di pubblicazione.

Nella logica del ragionamento che si sta cercando di sviluppare, è importante capire se all'interno dei plessi regolativi coinvolti nella costruzione del sistema integrato di fonti – anzitutto GDPR e *AI Act* – il paradigma dell'approccio basato sul rischio conviva con quello dei diritti fondamentali e se, in questa convivenza, trovi una declinazione allineabile allo standard tracciato dal quadro normativo prevenzionistico.

Rispetto al primo profilo, si può evidenziare come il Reg. 2024/1689 apra il proprio articolato normativo individuando, tra i propri obiettivi, la garanzia di «un livello elevato di protezione della salute, della sicurezza e dei diritti fondamentali sanciti dalla Carta dei diritti fondamentali dell'Unione europea» (art. 1, par. 1). Come segnala la dottrina, questo riferimento, in connessione alla «vocazione generale» e al «carattere orizzontale» dell'*AI Act*, implica un «passaggio logico indispensabile» per definire le regole da adottare ai fini della tutela; la protezione delineata dal Regolamento, infatti, «va ad affiancarsi […] a quella derivante in via generale dalla Carta, nonché dai principi enunciati nella giurisprudenza della Corte di giustizia»[13].

Quanto al GDPR, è importante sottolineare che se, ai sensi dell'art. 1, esso predispone regole in materia di protezione delle persone fisiche con riguardo al trattamento dei dati personali (par. 1), d'altra parte, nella medesima disposizione, si precisa che tale protezione attiene ai diritti e alle libertà fondamentali degli interessati (par. 2). Come chiarito dal cons. n. 4, il riferimento è a tutti i diritti fondamentali, alle libertà e ai principi riconosciuti dalla Carta dei diritti fondamentali dell'Unione europea e dai trattati. La giurisprudenza della Corte di giustizia, elaborata con riguardo al GDPR, conferma l'importanza della Carta nell'individuazione della fonte dei diritti fondamentali contemplati in tale quadro normativo e, in questa prospettiva, evidenzia l'obiettivo del Regolamento di garantire «un elevato grado di protezione dei diritti e delle libertà fondamentali delle persone fisiche»[14]. L'attenzione è, d'altra parte, per lo più concentrata sul bilanciamento tra il diritto al rispetto della vita privata e alla protezione dei dati personali e il diritto alla libertà di espressione e di informazione, di cui all'art. 17 GDPR e art. 11 della Carta[15],

[13] A. ADINOLFI, *L'intelligenza artificiale tra rischi di violazione dei diritti fondamentali e sostegno della loro promozione: considerazioni sulla (difficile) costruzione di un quadro normativo dell'Unione*, in A. PAJNO-F. DONATI-A. PERRUCCI (a cura di), *Intelligenza artificiale e diritto: una rivoluzione? Diritti fondamentali, dati personali e regolazione*, vol. I, il Mulino, Bologna, 2022, pp. 147-148.

[14] Cfr. da ultimo Corte giust., 4 ottobre 2024, *Patērētāju tiesību aizsardzības centrs*, C-507/23, ECLI:EU:C:2024:854, in cui si evidenzia la necessità di adottare una interpretazione delle disposizioni del GDPR «idonea a garantire la protezione dei dati personali come diritto fondamentale sancito all'art. 8, par. 1, della Carta, al quale rinvia il considerando 1 del GDPR» (par. 28). Cfr. in tal senso anche Corte giust. 4 ottobre 2024, *Koninklijke Nederlandse Lawn Tennisbond*, C-621/22, ECLI:EU:C:2024:858; Corte giust., 14 marzo 2024, *Újpesti Polgármesteri Hivatal*, C-46/23, ECLI:EU:C:2024:239; Corte giust., 7 marzo 2024, *IAB Europe*, C-604/22, ECLI:EU:C:2024:214.

[15] Cfr. Corte giust., 4 ottobre 2024, *Lindenapotheke*, C-21/23, ECLI:EU:C:2024:846; Corte giust., 8 dicembre 2022, *Google*, C-460/20, ECLI:EU:C:2022:962; Corte giust., 24 settembre 2019, *GC and Others v. CNIL*, C-507/17, ECLI:EU:C:2019:772, ove si chiarisce che «l'art. 52, par. 1, della Carta ammette che possano essere apportate limitazioni all'esercizio di diritti come quelli sanciti dagli articoli 7 e 8 della medesima, purché tali limitazioni siano previste dalla legge, rispettino il contenuto essenziale di detti diritti e libertà e, nel rispetto del principio di proporzionalità, siano necessarie e rispondano effettivamente a finalità di interesse generale riconosciute dall'Unione o all'esigenza di

ovvero il diritto a un ricorso effettivo, sancito dall'art. 47 della Carta[16], ovvero ancora il diritto di accesso del pubblico a documenti ufficiali[17].

Del resto, ai sensi dell'art. 51 della Carta, le disposizioni della stessa si applicano agli Stati membri (esclusivamente) nell'attuazione del diritto dell'Unione e la Corte di giustizia ha avuto modo di ribadire, da un lato, che «i diritti fondamentali garantiti nell'ordinamento giuridico dell'Unione si applicano in tutte le situazioni disciplinate dal diritto dell'Unione»[18], dall'altro, che «la definizione dell'ambito di applicazione *ratione materiae* del GDPR, quale enunciata all'articolo 2, par. 1, è molto ampia»: le eccezioni, previste dall'art. 2, par. 2 (ad es. «trattamenti effettuati per attività che non rientrano nell'ambito di applicazione del diritto dell'Unione», art. 2, par., 2 lett. a), «devono essere interpretate restrittivamente»[19].

A questo si aggiunga che, trattandosi di una normativa contenuta in un Regolamento, quindi già di per sé dotata di piena efficacia diretta, all'interno della sua interpretazione e applicazione, il riferimento alla Carta, come fonte del diritto fondamentale richiamato, non assume quella possibile portata e funzionalità che si riscontra, invece, laddove si tratti di una direttiva: in quest'ultimo caso, come dimostra il diritto antidiscriminatorio, il rinvio alla fonte primaria (nel caso specifico l'art. 21 della Carta) può sia costituire «una leva di Archimede»[20], in grado di attribuire un'efficacia diretta orizzontale alla regola di tutela (si pensi, ad esempio, a *Kücükdeveci*), sia incontrare, proprio per questo, una maggiore cautela (v. la pronuncia *J.K.*, dove si conferma l'applicazione della dir. 2000/78/CE anche all'accesso al lavoro autonomo, ma non viene mai richiamato l'art. 21 della Carta)[21].

proteggere i diritti e le libertà altrui» (punto 58). «Il regolamento 2016/679, e in particolare l'art. 17, par. 3, lett. a, prevede quindi espressamente il requisito del bilanciamento tra, da un lato, i diritti fondamentali al rispetto della vita privata e alla protezione dei dati personali, sanciti agli articoli 7 e 8 della Carta e, d'altro lato, il diritto fondamentale alla libertà di informazione, garantito dall'articolo 11 della Carta» (punto 59); «pertanto, il gestore di un motore di ricerca, quando riceve una richiesta di deindicizzazione [...], deve – sulla base di tutte le circostanze pertinenti della fattispecie e tenuto conto della gravità dell'ingerenza nei diritti fondamentali della persona interessata al rispetto della vita privata e alla protezione dei dati personali, sanciti dagli articoli 7 e 8 della Carta – verificare, alla luce [...] dell'art. 9, par. 2, lett. g, del regolamento 2016/679 e nel rispetto delle condizioni previste in tali disposizioni, se l'inserimento di detto link nell'elenco dei risultati, visualizzato in esito ad una ricerca effettuata a partire dal nome della persona in questione, si riveli strettamente necessario per proteggere la libertà di informazione degli utenti [...] sancita all'articolo 11 della Carta» (punto 68). l'ingerenza nei diritti fondamentali della persona interessata al rispetto della vita privata e alla protezione dei dati personali, sanciti dagli articoli 7 e 8 della Carta – verificare, alla luce [...] dell'art. 9, par. 2, lett. g, del regolamento 2016/679 e nel rispetto delle condizioni previste in tali disposizioni, se l'inserimento di detto link nell'elenco dei risultati, visualizzato in esito ad una ricerca effettuata a partire dal nome della persona in questione, si riveli strettamente necessario per proteggere la libertà di informazione degli utenti [...] sancita all'articolo 11 della Carta» (punto 68).

[16] Cfr. Corte giust., 2 marzo 2023, *Norra Stockholm Bygg AB*, C-268/21, ECLI:EU:C:2023:145.

[17] Cfr. Corte giust., 7 marzo 2024, *Endemol Shine Finland Oy*, C-740/22, ECLI:EU:C:2024:216.

[18] Corte giust., 26 febbraio 2013, Fransson, C-617/10, ECLI:EU:C:2013:105, punto 19.

[19] Cfr. Corte giust., 30 marzo 2023, *Hauptpersonalrat der Lehrerinnen und Lehrer*, C-34/21, ECLI: ECLI:EU:C:2023:270.

[20] M. BARBERA, *Il nuovo diritto antidiscriminatorio: innovazione e continuità*, in M. BARBERA (a cura di), *Il nuovo diritto antidiscriminatorio*, Giuffré, Milano, p. XIX ss., spec. p. XLV.

[21] Corte giust., 19 gennaio 2010, *Kücükdeveci*, C-555/07, ECLI: ECLI:EU:C:2010:21; Corte giust., 12 gennaio 2023, *J.K.*, C-356/21, ECLI:EU:C:2023:9.

Risulta, al riguardo, particolarmente significativa la segnalazione di chi osserva che al cuore del GDPR «sta la persona fisica con i suoi diritti e le sue libertà. La tutela dei dati personali è, si potrebbe dire, il contenuto "specialistico" di questa normativa che però, nelle sue caratteristiche fondamentali, è, per così dire, una "legge di rango costituzionale" di diritto europeo. [...] In sostanza, in questa prospettiva, la tutela dei dati personali costituisce, a partire dal livello europeo, l'attuazione di un diritto di spessore costituzionale, funzionale alla tutela "integrale" della persona e dei suoi diritti di libertà»[22].

Se le considerazioni svolte confermano l'importanza della grammatica dei diritti fondamentali nel contesto dei due principali plessi regolativi coinvolti nell'integrazione tra fonti, la riflessione richiede d'altra parte, come anticipato, anche un secondo passaggio, ossia una verifica della declinazione del paradigma procedurale *risk-based* all'interno di tali sistemi, nel confronto con lo standard delineato dal quadro prevenzionistico.

Tale verifica non può che partire dall'osservazione che le fonti coinvolte non solo non presentano una matrice lavoristica, ma rispondono anche a logiche diverse, anzitutto, nel caso dell'*AI Act*, a quella propria di una normativa di sicurezza di prodotto.

A tal proposito, il rischio, segnalato in dottrina, è che, da un lato, lo spostamento del baricentro sulla figura del fornitore/fabbricante che si riscontra all'interno di tale tipo di normativa marginalizzi il ruolo del datore-utilizzatore[23]; dall'altro che l'IA sia trattata «"solo" come un prodotto con un potenziale rischioso da contenere o eliminare», quando le sue implicazioni per le stesse «radici della materia» del lavoro sono molto più complesse[24]. Altro pericolo, sottolineato questa volta con particolare riguardo al raccordo con il GDPR, è che il grado di tutela previsto dalle fonti di matrice non lavoristica si fermi a una mera mitigazione del rischio, senza richiederne una eliminazione[25].

Rispetto al primo rilievo, si può anzitutto evidenziare come l'interazione con la normativa di prodotto non sia affatto inedita per il diritto del lavoro. È ancora una volta il quadro prevenzionistico a fornirne l'esempio, nel dialogo con la regolamentazione in materia di macchine, ora anche "intelligenti" (Reg. Ue 2023/1230, che entrerà in applicazione a partire da gennaio 2027; dir. 2006/42/CE, implementata a livello interno dal d.lgs. n. 17/2010). Quello che si può trarre, in sintesi, da questo raccordo, soprattutto per come approfondito dalla giurisprudenza, è che il sistema di prescrizioni indicate a carico del

[22] F. PIZZETTI, *La protezione dei dati personali e la sfida dell'Intelligenza Artificiale*, in F. PIZZETTI (a cura di), *Intelligenza artificiale, protezione dei dati personali e regolazione*, Giappichelli, Torino, 2018, p. 165.

[23] Cfr. L. ZOPPOLI, *Il diritto del lavoro dopo l'avvento dell'intelligenza artificiale: aggiornamento o stravolgimento? Qualche (utile) appunto*, in *WP CSDLE "Massimo D'Antona".IT*, 2024, n. 489; S. CIUCCIOVINO, *Risorse umane e intelligenza artificiale alla luce del regolamento (UE) 2024/1689, tra norme legali, etica e codici di condotta*, in *DRI*, 3, 2024, p. 573; T. TREU, *Regolamento su intelligenza artificiale e direttiva Due Diligence: a chi compete la valutazione dei rischi?*, in *Norme & Tributi Plus*, n. 37, 4 ottobre 2024.

[24] L. ZOPPOLI, *Il diritto del lavoro*…, cit., p. 14.

[25] G. GAUDIO, *Valutazioni di impatto e management algoritmico*, in *RGL*, 2024, n. 4, in corso di pubblicazione.

fabbricante non sposta il baricentro delle responsabilità del datore di lavoro. Certamente i requisiti di sicurezza che il fabbricante deve garantire[26] sono funzionali a consentire al datore di adempiere ai propri obblighi. Ad esempio, il fabbricante deve assicurare che i sistemi di controllo di macchine dotate di comportamenti auto-evolutivi siano in grado di permettere in qualsiasi momento la correzione della macchina per preservarne la sicurezza ovvero, in caso di macchina mobile autonoma, che le funzioni di supervisione consentano all'operatore di ricevere informazioni da remoto, anche sul verificarsi di situazioni impreviste o pericolose presenti o imminenti, che richiedono il suo intervento. D'altra parte, la posizione di garanzia di cui il datore è gravato in forza della normativa prevenzionistica concorre con quella del costruttore[27] e non è ad essa subordinata. E questo perché «la prossimità dell'imprenditore-datore alla fonte dei rischi, alle concrete modalità di lavoro e di eventuale elusione dei sistemi di sicurezza, gli consente immediatamente di percepire l'esposizione al pericolo dei lavoratori impiegati nell'utilizzo dei macchinari»[28]. «Il datore di lavoro ha l'obbligo di verificare la sicurezza delle macchine introdotte nella propria azienda e di rimuovere le fonti di pericolo per i lavoratori addetti all'uso di una macchina, a meno che questa non presenti un vizio occulto»[29], ossia un elemento di pericolo impossibile da accertare «per le speciali caratteristiche della macchina o del vizio di progettazione, che non consentano di apprezzarne la sussistenza con l'ordinaria diligenza»[30].

Il secondo rilievo segnalato richiede, invece, di verificare, all'interno dei plessi regolativi coinvolti, la possibilità o meno di allineare le diverse declinazioni dell'approccio *risk-based* alla logica prevenzionistica e alla conseguente specifica sequenza di obiettivi, tracciate dal diritto della sicurezza durante il lavoro.

Rispetto alla disciplina relativa al trattamento dei dati personali, è importante ricordare come la valutazione d'impatto di cui all'art. 35 GDPR, c.d. DPIA, preveda, in primo luogo, «una valutazione della necessità e della proporzionalità del trattamento in relazione alle finalità» (art. 35, par. 7, lett. b). Si è al riguardo segnalato come tale verifica imponga al datore di giustificare la scelta di un certo tipo di algoritmo

[26] La scansione di obiettivi individuata dalla dir. 89/391 per declinare l'approccio basato sul rischio è confermata nel Regolamento macchine, che al fine di garantire «un livello elevato di tutela della salute e sicurezza della persone» (art. 1), parimenti richiede al fabbricante di valutare il rischio e di progettare e costruire la macchina per eliminare i rischi o, ove non sia possibile, ridurre al minimo i rischi tenendo conto dei risultati della valutazione del rischio (All. III, parte B).

[27] Con riguardo all'applicabilità, nei confronti del fabbricante, del sistema della *product liability*, v. ora la nuova dir. Ue 2024/2853. La violazione della normativa di prodotto comporta d'altra parte, per il fabbricante, anche responsabilità di carattere penale. Ferma la configurabilità di più gravi fattispecie di reato in caso di lesioni o morte del lavoratore causate dal vizio di fabbricazione, ai sensi dell'art. 57, comma 2, d.lgs. 81/008, i fabbricanti e i fornitori che violano il disposto dell'art. 23, che vieta la fabbricazione di attrezzature di lavoro e impianti non rispondenti alle disposizioni legislative e regolamentari vigenti in materia di salute e sicurezza sul lavoro, sono puniti con l'arresto o l'ammenda.

[28] Cass pen., sez. IV, 6 aprile 2011, n. 33285.

[29] Cass. pen., sez. IV, 30 settembre 2016, n. 44327; Cass. pen., sez. IV, 30 maggio 2018, n. 29144; Cass. pen., sez. IV, 26 gennaio 2021 n. 5794; Cass. pen., sez. IV, 22 settembre 2021, n. 36153; Cass. pen., Sez. IV, 13 marzo 2023, n. 10398; Cass. pen., Sez. IV, 3 aprile 2024, n. 13382.

[30] Cass. pen., sez. IV, 27 ottobre 2021, n. 41147. Cfr. Cass. pen., 30 maggio 2013, n. 26247, con nota di I. SCORDAMAGLIA, *Malfunzionamento delle macchine e delle attrezzature di lavoro: le concorrenti responsabilità penali del datore di lavoro, del fabbricante e del fornitore*, in *Cass. pen.*, 2014, n. 4, p. 1340 ss.

rispetto a un determinato obiettivo, a fronte dell'assenza di alternative meno invasive per i diritti e le libertà coinvolti[31]. Si può ritenere imponga, in tal senso, una valutazione analitica del processo organizzativo nell'ottica dell'eliminazione del rischio e una giustificazione della scelta del tipo di trattamento (e di algoritmo) a fronte dell'assenza di alternative meno invasive per i diritti e le libertà coinvolti. Solo in via subordinata, ossia laddove tali alternative manchino e il rischio non si possa pertanto eliminare, la questione si sposta sull'individuazione di misure adeguate a mitigare, ridurre i rischi per *tutti* i diritti fondamentali dei lavoratori (nell'ottica di un elevato grado di tutela, v. *supra*) e garantire il rispetto di tutti i principi di protezione del GDPR, tra cui quello di minimizzazione dei dati (e i corollari principi di necessità, pertinenza e non eccedenza)[32]. In questa lettura del disposto normativo dell'art. 35 si rafforzerebbe, quindi, l'allineamento anche del modello *risk-based* adottato nel GDPR con lo standard tracciato dal quadro in materia di salute e sicurezza. La similarità delle procedure e di un approccio metodologico parimenti ispirato alla cultura prevenzionistica è tanto più valorizzata da quella parte della dottrina che, muovendo da tale premessa, riflette sulla possibilità di attrarre la tutela dei dati personali in quella della salute ampiamente intesa. In tal senso, è promossa una sistematizzazione concettuale del profilo della sicurezza dei dati «nella portata espansiva dell'art. 2087 c.c.» e la configurabilità di «un documento di valutazione per la prevenzione del rischio *privacy* […] come *addendum* del consueto documento di valutazione dei rischi sul lavoro»[33].

L'allineamento esaminato con riguardo al GDPR riesce a trovare conferma anche rispetto all'*AI Act*. È sì vero, infatti, che, in base all'art. 9, par. 3, i rischi interessati dalle misure di gestione che il fornitore deve predisporre sono «solo quelli che possono essere ragionevolmente attenuati o eliminati attraverso lo sviluppo o la progettazione del sistema di IA ad alto rischio o la fornitura di informazioni tecniche adeguate» (art. 9, par 3). È importante evidenziare, tuttavia, come il quinto paragrafo dell'art. 9, nel riferirsi agli obiettivi di eliminazione e attenuazione del rischio, riprenda l'ordine presente nella direttiva su salute e sicurezza. Nell'individuare le misure di gestione dei rischi più appropriate, il fornitore deve, infatti, garantire: «l'eliminazione o la riduzione dei rischi individuati e valutati» e, «ove opportuno, l'attuazione di adeguate misure di attenuazione e di controllo nell'affrontare i rischi che non possono essere eliminati».

[31] H. JANSSEN, M. SENG AH LEE, J. SINGH, *Practical fundamental rights impact assessments*, in *International Journal of Law and Information Technology*, 30, 2022, p. 200 ss. La sequenza "eliminazione – mitigazione" del rischio, nel contesto della DPIA, è valorizzata anche da A. MANTELERO, *Art. 35*, in R. D'ORAZIO, G. FINOCCHIARO, O. POLLICINO, G. RESTA (a cura di), *Codice della privacy e data protection*, Giuffré, 2021, pp. 532 ss.

[32] Ai sensi dell'art. 5, par. 1, lett. c, i dati devono essere adeguati, pertinenti e limitati a quanto necessario rispetto alle finalità per le quali sono trattati. Come chiarisce il cons. n. 39, «i dati personali dovrebbero essere trattati solo se la finalità del trattamento non è ragionevolmente conseguibile con altri mezzi». Sulla rilevanza che possono assumere i principi di limitazione della finalità e di minimizzazione in caso di utilizzo di *big data* cfr. Corte giust. 4 ottobre 2024, *Meta Platforms Ireland Ltd*, C-446/21, ECLI:EU:C:2024:834.

[33] L. D'ARCANGELO, *La tutela del lavoratore nel trattamento dei dati personali*, Aracne, 2024, pp. 168 e 110

All'interno dell'*AI Act*, l'approccio arriva a proiettarsi anche sulla figura del *deployer*, obbligato a utilizzare la macchina in conformità alle istruzioni consegnate dal fornitore, quindi dando anche applicazione alle misure di sorveglianza umana indicate, finalizzate espressamente, a norma dell'art. 14, in sequenza, a prevenire o ridurre al minimo i rischi per la salute e sicurezza o i diritti fondamentali. Altrettanto significativo è che il *deployer* sia chiamato a monitorare debitamente il funzionamento del sistema (anche attraverso la registrazione e conservazione dei *log* di tracciamento) e debba sospenderne l'uso, qualora abbia ragione di ritenere, a fronte del monitoraggio, che nonostante il rispetto delle istruzioni ricevute il sistema presenti un rischio per la salute e sicurezza e i diritti fondamentali.

1.1. Alla ricerca di un obbligo trasversale di valutazione di impatto in un sistema di fonti a geometria variabile

Ferme le considerazioni svolte nel paragrafo precedente, rimane un dato critico: l'assenza, a livello normativo, di un obbligo datoriale di valutazione del rischio derivante dall'uso di sistemi automatizzati, di carattere generale e trasversale, che prescinda dalla natura del dato trattato ovvero dal tipo di bene esposto al rischio di lesione. L'obbligo di effettuare la DPIA, per quanto inevitabile in caso di sistemi automatizzati[34], segue l'ambito di applicazione del GDPR e, quindi, il raggio d'azione consentito dalla pur ampia nozione di dato personale[35] (con inclusione dei dati non personali, se indissolubilmente legati a quelli personali all'interno del dataset[36]). L'obbligazione di sicurezza, anche nella sua declinazione procedurale di valutazione e gestione del rischio, concerne (solo) la tutela della salute, per quanto ampiamente intesa, nonché affiancata dalla dignità nella norma generale e di chiusura fornita a livello interno dall'art. 2087 c.c.[37].

In tal senso, emerge con forza il limite del compromesso raggiunto in seno all'*AI Act* rispetto alla previsione dell'obbligo di valutazione di impatto sui diritti fondamentali di cui all'art. 27, un limite che

[34] Si consenta di rinviare sul punto a M. PERUZZI, *op cit.*

[35] A fronte dell'ampia definizione prevista dall'art. 4, par. 1, n. 1, personali sono, ad esempio, anche i dati relativi ai movimenti di uno "scheletro digitale" e alla sua interazione con un co-bot ovvero quelli concernenti l'uso di una macchina o di uno strumento di lavoro (dalla geolocalizzazione alle battute di tastiera), se ricollegabili a un lavoratore, identificabile anche solo sulla base dell'incrocio con i dati dei turni di lavoro, e utilizzabili per valutare la sua produttività o le sue prestazioni o comunque per adottare misure incidenti sulle sue condizioni di lavoro. Si consideri, inoltre, che il carattere anonimo del dato, tale da escluderlo dall'ambito di applicazione del GDPR, dipende dall'impossibilità di recuperare il collegamento con l'interessato, tenuto conto di tutti i mezzi di cui il titolare del trattamento o un terzo possano ragionevolmente avvalersi, anche a fronte dell'evoluzione tecnologica (cons. n. 26). Oltre a porsi, quindi, come elemento dinamico, esso è reso sempre più complesso dalle potenzialità della digitalizzazione e dall'interconnettività dei dispositivi nell'*Internet of Things* (cfr. sul punto Cfr. P. TULLINI, *Dati*, in M. NOVELLA-P. TULLINI, *Lavoro digitale*, Giappichelli, Torino, 2022, pp. 105 ss.; nonché da ultimo G. FINOCCHIARO, *Diritto dell'intelligenza artificiale*, Zanichelli, Bologna, 2024, in cui si evidenzia che «la qualificazione di dato anonimo [...] dipende fortemente dalle risorse, soprattutto tecnologiche, disponibili», p. 64).

[36] In tal senso l'art. 2, par. 2, Regolamento (Ue) 2018/1807 in materia di libera circolazione dei dati non personali nell'Unione europea. Il profilo è rimarcato in più punti anche dal Regolamento (Ue) 2023/2854 (c.d. *Data Act*).

[37] Rispetto alla portata espansiva dell'art. 2087 c.c. v. la riflessione di L. D'ARCANGELO, *op. cit.*, di cui *supra*.

attiene sia all'ambito di applicazione dell'obbligo, sia alla concezione dello strumento, tradita dai termini con cui è predisposto il processo.

L'obbligo, proposto dal Pe, è stato infatti circoscritto, nell'accordo interistituzionale raggiunto e quindi nel testo finale, agli organismi di diritto pubblico o agli enti privati che forniscono servizi pubblici, come ospedali, scuole, banche e compagnie di assicurazione. Ciò pare suggerire un'attenzione specifica ai diritti fondamentali dell'utenza, posto che la rilevanza dei diritti dei lavoratori, pur indicati tra i diritti fondamentali oggetto di tutela nel contesto dell'*AI Act* (v. cons. n. 48), avrebbe necessariamente richiesto un'applicazione trasversale dello strumento[38].

Dal punto di vista dei contenuti, alla luce della valutazione, i *deployer* dovrebbero «stabilire le misure da adottare al concretizzarsi dei rischi individuati, compresi, ad esempio, i meccanismi di *governance* […], quali le modalità di sorveglianza umana secondo le istruzioni per l'uso, o le procedure di gestione dei reclami e di ricorso, dato che potrebbero essere determinanti nell'attenuare i rischi per i diritti fondamentali in casi d'uso concreti» (cons. n. 96).

A sottolineare lo iato tra l'obbligo in esame e quello di valutazione dei rischi o di impatto previsto in altri sistemi normativi (come quello prevenzionistico e di protezione dei dati personali) sono, d'altra parte, le modalità di esecuzione stabilite dal paragrafo 5 dell'art. 27: si prevede, infatti, che l'obbligo debba essere adempiuto attraverso la compilazione di un modello di questionario predisposto dall'Ufficio per l'IA, dichiaratamente volto ad agevolare la semplificazione del processo. Il modello compilato deve essere presentato all'autorità di vigilanza del mercato nell'ambito della notifica dei risultati della valutazione[39].

Non presenta questi limiti e, anzi, include un approccio partecipato[40] l'obbligo di valutazione di impatto previsto dalla direttiva (Ue) 2024/2831 in materia di lavoro mediante piattaforme digitali. Il limite, in questo caso, risiede nell'ambito di applicazione dello strumento, circoscritto allo specifico settore interessato dal quadro normativo.

L'art. 10, nel contesto delle garanzie di sorveglianza umana, richiede in particolare che le piattaforme, con la partecipazione dei rappresentanti dei lavoratori, effettuino regolarmente, e in ogni caso ogni due anni,

---

[38] Cfr. A. ALAIMO, *Il Regolamento sull'Intelligenza Artificiale. Un treno al traguardo con alcuni vagoni rimasti fermi*, in questa rivista, 2024, n. 25, pp. 231 ss.

[39] Come segnalato da A. MANTELERO, «The questionnaire to be developed by the AI Office under Article 27 (5) may […] assist deployers in fulfilling certain obligations of the FRIA, particularly in the planning and scoping phases, as well as some aspects of data collection during the assessment phase. However, a purely questionnaire-based approach, and even worse its automation (which requires a high degree of uniformity), cannot fully capture the contextual nature of the FRIA and needs to be integrated with a methodology for risk quantification and management» (*The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, legal obligations and key elements for a model template*, in *Computer Law and Security Review*, n. 54, 2024, p. 9).

[40] Si rinvia ad altra sede la riflessione sulla declinazione della dimensione collettiva dell'approccio *risk-based*. Cfr. *ex multis* L. ZAPPALÀ, *Sistemi di IA ad alto rischio e ruolo del sindacato alla prova del risk-based approach*, in *LLI*, 2024, n. 10(1), p. 52 ss., nonché, con riguardo all'analisi del profilo nel contesto della direttiva piattaforme, M. AIMO, *Trasparenza algoritmica nel lavoro su piattaforma: quali spazi per i diritti collettivi nella proposta di Direttiva in discussione?*, in *LDE*, 2024, n. 2;

una valutazione dell'impatto delle decisioni individuali prese o sostenute dai sistemi automatizzati, con conseguente adozione delle misure necessarie, compresa, se del caso, una modifica del sistema o la cessazione del suo utilizzo. Il processo deve essere svolto da parte di specifiche persone incaricate e dotate della competenza, formazione e autorità necessarie per esercitare tale funzione (comprese quelle necessarie per non accogliere le decisioni automatizzate e non cadere, quindi, nei c.d. *automation bias*[41]). Le informazioni sui risultati sono trasmesse ai rappresentanti dei lavoratori, nonché messe a disposizione dei lavoratori e delle autorità nazionali competenti su loro richiesta.

La valutazione concerne tutti i processi automatizzati, a prescindere dalla natura del dato trattato (e infatti lo strumento previsto dall'art. 10 si affianca alla DPIA, comunque oggetto di disciplina rafforzata nella logica di tutela del lavoratore, anche collettiva, a norma dell'art. 8), e l'impatto da valutare riguarda tutte le condizioni di lavoro, compresa espressamente la parità di trattamento (sull'importanza dell'approccio *risk-based* rispetto alla discriminazione algoritmica, v. *infra*).

Il profilo è centrale e non è un caso che l'Etuc abbia ribadito a più riprese, da ultimo a ottobre 2024 in occasione della settimana europea sulla salute e sicurezza, la necessità di un intervento legislativo Ue che estenda a tutti gli ambiti lavorativi la tutela contro la gestione algoritmica prevista per i lavoratori su piattaforma[42].

In questa prospettiva, può essere significativo riflettere sul ruolo che potrà assumere la dir. Ue 2024/1760 sulla c.d. *due diligence* a sostegno di una trasfusione trasversale dell'approccio *risk-based* nell'ambito delle regole di tutela del lavoro, anche quindi rispetto ai rischi derivanti dall'IA. Il «dovere di diligenza» ivi previsto si configura, infatti, come un obbligo di valutazione e gestione degli impatti delle attività dell'impresa per una serie di diritti fondamentali, senza pregiudizio per i titoli di responsabilità derivanti dalla violazione delle specifiche fonti di tutela degli stessi[43]. Come evidenziato dalla dottrina, la garanzia di esercizio e integrazione del «dovere di diligenza basato sul rischio» in tutte le politiche della società (art. 5), comprese quelle del personale, si traduce nell'obbligo datoriale di valutare i rischi delle proprie attività

[41] Si noti come tale profilo sia già espressamente valorizzato dal Garante della privacy nell'interpretazione dell'art. 22 GDPR. Nel provvedimento del 13 novembre 2024 adottato nei confronti di Foodinho con riguardo all'utilizzo del sistema automatizzato di gestione dell'attività dei rider, il garante ha ordinato alla società di conformare al Regolamento i propri trattamenti, garantendo all'interessato, tra le varie misure, il diritto di ottenere l'intervento umano e assicurando a tal fine «un'adeguata formazione degli operatori addetti nonché la possibilità per gli operatori stessi di ignorare, se del caso, l'*output* del processo algoritmico, per evitare la possibile tendenza a farvi automaticamente affidamento».

[42] V. ETUC, *Directives needed to make work safe for digital age*, 23 ottobre 2024, etuc.org. La prospettiva, già promossa dall'Etuc in una risoluzione del 6 dicembre 2022, era stata condivisa sia da Gualmini e Benifei all'interno del Parlamento europeo, sia da Schmit, commissario europeo per il lavoro e i diritti sociali nella Commissione von der Leyen I.

[43] Ai sensi dell'art 29, par. 6, le norme in materia di responsabilità civile previste dalla direttiva «non limitano la responsabilità delle società ai sensi dei sistemi giuridici dell'Unione o nazionali e lasciano impregiudicate le norme unionali o nazionali in materia di responsabilità civile relative agli impatti negativi sui diritti umani o agli impatti ambientali negativi che prevedono la responsabilità in situazioni non contemplate dalla presente direttiva o che prevedono una responsabilità più rigorosa».

anche «per prevenire e mitigare lesioni ai diritti fondamentali dei lavoratori»[44]. In tale prospettiva, si evidenziano l'importanza e la funzionalità del paradigma procedurale *risk-based*: per quanto la direttiva, infatti, faccia riferimento a «una base di diritti da tempo riconosciuti nei paesi membri compresa l'Italia», d'altra parte «il tasso di effettività delle normative di tutela non è sempre garantito specie […] su alcuni aspetti trasversali, come il divieto di disparità di trattamento in materia di lavoro e di retribuzioni (allegato alla direttiva, parte I, n. 14)»; in tale ottica, le indicazioni della direttiva nella traduzione del «dovere di diligenza» si presentano «potenzialmente rilevanti per rafforzare la effettiva osservanza dei diritti dei lavoratori», anche nel contesto inciso dall'uso dell'IA[45]. In collegamento con il ragionamento sviluppato nel corso dell'analisi, può essere interessante segnalare come la direttiva articoli il processo di gestione dei rischi nella seguente sequenza di fasi: «prevenzione e attenuazione degli impatti negativi potenziali e arresto degli impatti negativi effettivi e minimizzazione della relativa entità» (art. 5, par. 1, lett. c; v. anche cons. n. 20). Emerge, cioè, anche qui la priorità per la prevenzione degli impatti potenziali e l'arresto immediato di quelli effettivi, rispetto all'opzione subordinata della mitigazione e minimizzazione, che pur «dovrebbe comportare un esito il più possibile vicino all'arresto»[46]. Rimane anche qui, d'altra parte, come dato critico, il limite del campo di applicazione dell'obbligo, che incombe solo sulle società caratterizzate da determinate dimensioni (art. 2), per quanto con riguardo sia alle loro attività, sia a quelle delle loro filiazioni, nonché dei loro partner commerciali se collegate alla loro catena di attività (art. 8, par. 1).

1.2. La necessità di integrare l'approccio *risk-based* per garantire l'effettività dei diritti in un contesto di utilizzo dell'IA: l'esempio della *proxy discrimination*. La prospettiva volontaristica

Si è più volte evidenziato come il paradigma procedurale dell'approccio basato sul rischio, sul duplice piano del governo dei dati e della valutazione di impatto dei processi decisionali, sia indispensabile per garantire l'effettività dei diritti in un contesto di utilizzo dell'IA.

[44] T. TREU, *op. cit.*. Rispetto alla base giuridica della direttiva, fondata sugli artt. 50 e 114 del TFUE, F. GUARRIELLO evidenzia come «la normativa in parola non possa essere letta, interpretata e applicata come mera disciplina di armonizzazione del diritto societario, pena una visione for- temente riduttiva degli obiettivi da essa perseguiti. Come per altri plessi normativi di recente emanazione (v. il regolamento sull'IA o il *Digital Service Act*), la base giuridica fondata sul funzionamento del mercato interno non ha impedito di allestire una disciplina ambiziosa, che travalica l'accezione strettamente funzionale al mercato, per fondarsi su un sistema assiologico che comprende tutti gli interessi in gioco e che pone al centro la dignità della persona e la protezione dell'ambiente e del clima come *habitat* naturale da tutelare e trasmettere intatto alle generazioni future» (Take Due Diligence Seriously*: commento alla direttiva 2024/1760*, in *DLRI*, n. 3, 2024, p. 252).

[45] T. TREU, *op. cit.*

[46] A norma dell'art. 11, par. 2, «Laddove l'arresto immediato dell'impatto negativo risulti impossibile, gli Stati membri provvedono a che le società ne minimizzino l'entità». Come chiarisce il cons. n. 53, «La minimizzazione dell'entità degli impatti negativi dovrebbe comportare un esito il più possibile vicino all'arresto dell'impatto negativo».

In tal senso, si è segnalato come dato critico centrale il fatto che a livello normativo non sia trasversalmente garantita la penetrazione di questo paradigma nelle posizioni d'obbligo poste in capo al datore(-*deployer*) dalle fonti di tutela del lavoratore.

È vero che, nell'ambito del Regolamento 2024/1689, le prescrizioni a carico del fornitore sono funzionali a prevenire i rischi e a supportare il datore-*deployer* nell'adempimento dei propri obblighi (si pensi alle misure volte a garantire la sorveglianza umana, nonché la comprensibilità del funzionamento e degli *outcome* del sistema). Lo stesso *deployer* è gravato, inoltre, dall'obbligo di garantire la pertinenza e sufficiente rappresentatività dei dati di *input* (nella misura in cui esercita il controllo su di essi, art. 26, par. 4), nonché di monitorare costantemente il sistema, dovendo sospenderne l'uso, qualora rilevi un rischio per i diritti fondamentali, nonostante il rispetto delle istruzioni ricevute (art. 26, par. 5). In tal senso, si è ritenuto di poter evidenziare una proiezione del paradigma *risk-based* sulla figura del datore(-*deployer*), che pur non arriva però a perfezionare un obbligo generale di valutazione di impatto sui diritti fondamentali/sulle condizioni di lavoro, trasversalmente applicabile.

L'importanza di tale obbligo si evince in modo significativo nel confronto con l'applicazione del divieto di discriminazione nel contesto algoritmico. A conferma si può ricordare l'attenzione dedicata al profilo negli ambiti in cui un obbligo trasversale sussiste: sia nella direttiva piattaforme, sia nella riflessione dottrinale che valorizza, nella prospettiva dell'IA, il paradigma basato sul rischio della *due diligence* rispetto allo scarso tasso di effettività di alcuni diritti fondamentali (v. *supra*).

La necessità di introdurre nella posizione d'obbligo datoriale il paradigma procedurale dell'approccio *risk-based* si spiega in ragione della complessità che connota il fenomeno della discriminazione algoritmica, in particolare la c.d. *proxy discrimination*.

Per comprendere le peculiarità di quest'ultima fattispecie, si può partire da due considerazioni.

Anzitutto, il rilievo dei fattori di rischio a livello di gruppo determina il tendenziale carattere sistematico delle discriminazioni, rispecchiato nell'importanza assunta, nel diritto antidiscriminatorio, dalla prova statistica ai fini dell'offerta della prova, pur semipiena, sulla sussistenza di un nesso di causalità tra una disparità di trattamento subita e il fattore di rischio. In secondo luogo, ai fini della tutela garantita dal divieto di discriminazione (indiretta), a rilevare per la configurazione dell'impatto differenziato (e ingiustificato) sono elementi che risultano maggiormente ricorrenti all'interno del gruppo protetto, elementi che, in tal senso, arrivano a caratterizzarlo in modo particolare anche in termini di conseguente svantaggio subito.

I due profili evidenziati, ossia la sistematicità delle discriminazioni e la riconoscibilità o meno di *pattern* all'interno dei gruppi (protetti e non), si combinano nella *proxy discrimination*. Proprio gli elementi ricorrenti nel gruppo sottorappresentato nel dataset di addestramento diventano i "rappresentanti" del fattore di

rischio che caratterizza tale gruppo, usati dall'algoritmo per penalizzare i soggetti che vi appartengono. Specularmente, gli elementi ricorrenti nel gruppo sovrarappresentato sono usati per dare priorità allo stesso, in quanto statisticamente più rispondente a un determinato risultato (apprendimento supervisionato), ovvero sono utilizzati per individuare lo standard su cui rilevare anomalie o a partire dal quale ridurre la dimensionalità dei dati e individuare le componenti principali che possono migliorare la performance dell'algoritmo (apprendimento non supervisionato).

In questi casi, il tema della qualità dei dati diventa centrale, soprattutto se si considera che l'eliminazione del fattore protetto dai *training data* non esclude che il sistema lo inferisca implicitamente dalle correlazioni con altre caratteristiche, che diventano in tal senso *proxy* del fattore e la base per la disparità di trattamento di una categoria tutelata. È proprio la scarsa qualità dei dati sotto il profilo della loro rappresentatività a portare il sistema a "imparare" a penalizzare i gruppi sottorappresentati nel dataset di apprendimento, in quanto non rientranti tra quelli "statisticamente migliori" rispetto all'obiettivo dato. L'operazione è effettuata utilizzando le caratteristiche che il sistema trova associate a tali gruppi: il codice postale può diventare così il *proxy* per fattori come la razza, l'origine etnica, la religione o l'orientamento sessuale; il tipo di occupazione, l'utilizzo del part-time, le interruzioni di carriera ovvero anche il background scolastico e le attività extracurriculari possono costituire *proxy* per il genere; i nomi di battesimo per l'età[47]. Ciò che è interessante osservare è come in questi casi la fattispecie presenti una stretta prossimità alla tipologia della discriminazione diretta, poiché gli elementi che fungono da *proxy* del fattore di rischio, pur apparentemente neutri, sono individuati e usati dalla macchina proprio perché rappresentanti del fattore di rischio, a partire dal nesso che la macchina individua – in termini statistici[48] – tra il fattore di rischio (presente o inferito) e il risultato assegnato (l'etichetta, nell'apprendimento supervisionato) ovvero a partire dall'individuazione, all'interno di dati non etichettati, del gruppo privo del fattore come insieme

[47] Cfr. A. ER PRINCE-D. SCHWARCZ, *Proxy discrimination in the age of artificial intelligence and big data*, in *Iowa L. Rev.*, vol. 105, 2019, p. 1257 ss.; I. BARTOLETTI-R. XENIDIS, *Study on the impact of artificial intelligence systems, their potential for promoting equality, including gender equality, and the risks they may cause in relation to non-discrimination*, Council of Europe, August 2023; H. WEERTS-A. KELLY-LYTH-R. BINNS-J. ADAMS-PRASSL, *Unlawful Proxy Discrimination: A Framework for Challenging Inherently Discriminatory Algorithms*, in arXiv:2404.14050v1, 22 aprile 2024; X. CHEN, *Algorithmic proxy discrimination and its regulations*, in *Computer Law & Security Review*, vol. 54, 2024, p. 1 ss.; A. TOPO, *Nuove tecnologie e discriminazioni*, XXI Congresso Nazionale Aidlass, *Diritto antidiscriminatorio e trasformazioni del lavoro*, 2024, p. 24.

[48] Nel confronto dottrinale sulle categorie normative di discriminazione a cui sia, a seconda dei casi, ascrivibile la *proxy discrimination*, ovvero sulla necessità di introdurre una nozione ad hoc, M. BARBERA evidenzia come «anche il nesso di causalità può basarsi su una correlazione probabilistica, e non su una necessaria certezza, in virtù del carattere induttivo che può assumere la prova del nesso di causalità stesso» (*Discriminazioni algoritmiche e forme di discriminazione*, in *LLI*, 2021, n. 7(1), p. I.13). Sul punto, cfr. altresì R. Xenidis, *Tuning EU equality law to algorithmic discrimination: Three pathways to resilience,* in *Maastricht Law Journal of European and Comparative Law*, 2020, vol. 27, 736 ss.; G. DE SIMONE, *Discriminazione*, in M. NOVELLA, P. TULLINI (a cura di), Lavoro digitale, Giappichelli, Torino, p. 127 ss.; R. VOZA, *Eguaglianza e discriminazioni nel diritto del lavoro. Un profilo teorico*, FrancoAngeli, Milano, 2024.

dominante e *top-performing*, rispetto al quale distinguere *pattern* anomali o costruire *benchmark* di riferimento[49].

Le difficoltà che pongono l'intercettazione e il contrasto delle discriminazioni algoritmiche nel contesto del *machine learning* evidenziano con forza la necessità di raccordare le garanzie di trasparenza a quelle di sorveglianza umana e di assicurare tale sorveglianza sia sul piano della *governance* dei dati (e in tal senso assumono specifico rilievo alcuni obblighi previsti nell'*AI Act*[50] in capo al fornitore e al *deployer* per evitare la scarsa rappresentatività dei dati ovvero la presenza di *bias* nei dataset impiegati[51]), sia su quello degli effetti derivanti dall'utilizzo del sistema.

È una prospettiva che trova terreno fertile nel diritto antidiscriminatorio, vista l'introduzione, a opera della dir. 2023/970, della valutazione congiunta quale strumento (partecipato) di contrasto alla discriminazione retributiva di genere, fattispecie non a caso parimenti connotata da un elevato grado di problematicità sotto il profilo del tracciamento e correzione delle cause.

Tale strumento potrebbe fornire un modello di riferimento anche per il contrasto delle discriminazioni nel contesto dei sistemi automatizzati, da declinare nel senso di una valutazione periodica (e partecipata) del loro impatto sulle condizioni di lavoro, con specifico riguardo alla parità di trattamento dei lavoratori[52]. Proprio la previsione di indicatori sentinella, quali differenze di impatto non riconducibili a

[49] Nel contesto dell'apprendimento non supervisionato e a differenza dell'esempio da ultimo proposto, l'utilizzo di *proxy* per la creazione di *cluster*, ossia per l'identificazione di gruppi di lavoratori con caratteristiche simili, può essere, d'altra parte, maggiormente avvicinata alla discriminazione indiretta. Ciò nel momento in cui la scelta delle variabili, selezionate per raccogliere i dati dei lavoratori che la macchina dovrà analizzare in sede di apprendimento, sia effettuata con un approccio che è probabilmente orientato all'oggettività, ma non risulta fondato su elementi appropriati e necessari rispetto alla finalità per cui sono utilizzati i risultati del *clustering*. Lo stesso approccio, per capirci, che contraddistingue alcune prassi unilaterali o negoziali in azienda sui premi di produttività, che utilizzano la variabile della presenza in servizio come criterio di calcolo, con conseguente impatto differenziato a detrimento delle donne ovvero, più in generale, dei lavoratori con carichi di cura. La riflessione sulla

[50] Nell'ottica del diritto antidiscriminatorio è, altresì, importante ricordare che ai sensi dell'art. 77 dell'*AI Act* gli organismi per la parità, in quanto autorità chiamate a presidiare l'applicazione del diritto dell'Unione che tutela i diritti fondamentali, hanno accesso alla documentazione creata a norma del Regolamento, ove necessario per il loro mandato (v. cons. n. 157).

[51] Al fornitore si richiede, ad esempio, di garantire, attraverso adeguate pratiche di governance e gestione, dati di alta qualità (in rapporto alla finalità prevista e tenendo conto delle caratteristiche dell'ambito in cui il sistema di IA è destinato a essere usato, come l'ambiente di lavoro), in particolare in caso di tecniche che prevedono l'addestramento di modelli, così da assicurare che il sistema di IA ad alto rischio non diventi una fonte di discriminazione vietata dal diritto dell'Unione (art. 10; cons. n. 67). Sempre il fornitore è tenuto a inserire nella documentazione tecnica – oggetto di valutazione, interna o da parte di terzi, ai fini della verifica di conformità del sistema di IA, poi da conservare, aggiornare e tenere a disposizione delle autorità competenti – informazioni dettagliate sui prevedibili risultati indesiderati e fonti di rischio di discriminazione in considerazione della finalità prevista del sistema di IA (art. 11; allegato IV). I metodi per rilevare distorsioni identificabili rientrano, altresì, tra le informazioni che tutti i fornitori di modelli di GPIA devono includere nella documentazione tecnica (art. 53; allegato XI). Per quanto riguarda il *deployer*, il Regolamento richiede, come visto, che garantisca che i dati di input siano pertinenti e sufficientemente rappresentativi, nella misura in cui esercita il controllo sugli stessi, profilo centrale in caso di sistemi caratterizzati da adattabilità e apprendimento continuo (art. 26, par. 4).

[52] La prospettiva è inquadrata da S. CIUCCIOVINO in termini di «trasparenza di impatto», individuata come garanzia «che implica una valutazione degli effetti dell'impiego della IA, sia a livello preventivo mediante la valutazione e

criteri giustificati o legittimi, come nel caso della direttiva n. 970, potrebbe costituire un interessante punto di raccordo tra le garanzie di trasparenza e quelle di sorveglianza umana.

A tal proposito, può essere significativo richiamare, con uno sguardo oltreoceano, quanto stabilito dalla *New York City Local Law n. 144/2021*: tale legge richiede ai datori di lavoro o all'agenzie di collocamento, che utilizzano strumenti automatizzati per selezionare i candidati all'assunzione o a una promozione, di sottoporre il processo a una verifica dei *bias* da parte di un auditor indipendente e di pubblicarne i risultati sul proprio sito web. A essere pubblicati sono, segnatamente, il tasso di selezione/punteggio e l'indice di impatto (ossia il rapporto con il tasso di selezione/punteggio della categoria più selezionata) per le categorie di sesso (uomini e donne), per una serie di categorie di razza/etnia (ispanici o latini, bianchi, afroamericani, nativi hawaiani o delle isole del Pacifico, asiatici, nativi americani o dell'Alaska, due o più razze[53]), nonché per categorie intersezionali di sesso, etnia e razza.

La normativa non individua indici "sentinella", né specifiche conseguenze a fronte di determinati differenziali. Il modello, d'altra parte, non può che far riflettere, soprattutto se il tentativo, in una prospettiva di ricerca di strumenti per intercettare e contrastare le discriminazioni algoritmiche, è quello di agganciare alla misura di trasparenza un meccanismo di sorveglianza umana, come avviene nel contesto della valutazione congiunta di cui alla direttiva n. 970, ma – a differenza di quest'ultima – oltre il (mero) confronto di genere.

È un punto di riflessione, certamente delicato, che anche la direttiva piattaforme non incontra, nel momento in cui predispone, a tutela della parità di trattamento, il meritorio strumento della valutazione periodica di impatto[54].

Nel ragionamento sulla declinazione dell'approccio *risk-based* nell'ambito antidiscriminatorio, è altresì importante ricordare come la gestione del rischio non incida, in questo caso, sulla determinazione del criterio di imputazione soggettivo della responsabilità, essendo la sussistenza di colpa o dolo irrilevante ai fini della configurazione della fattispecie vietata (ma non anche ai fini del risarcimento del danno). Essa

---

ponderazione preventiva del rischio e la responsabilizzazione del titolare del trattamento, sia ex post, in esito all'applicazione del sistema per verificarne appunto gli impatti a livello individuale e di gruppo» (*op. cit.*, p. 597).

[53] Le categorie indicate sono quelle previste nel c.d. «*EEO-1 Component 1*» report, una relazione annuale sui dati demografici della forza lavoro, richiesta dalla *US Equal Employment Opportunity Commission* (EEOC) ai datori di lavoro con almeno 100 dipendenti o agli appaltatori federali con almeno 50 dipendenti, al ricorrere di determinati criteri.

[54] Si segnala come la disposizione in esame, a differenza di quella su sicurezza e salute (art. 12), trovi applicazione nei confronti di tutti i lavoratori su piattaforma, subordinati e genuinamente autonomi. La differenza di ambito soggettivo si spiega in ragione della duplica base giuridica della direttiva: l'art. 153, par. 1, lett. b, in materia di miglioramento delle condizioni di lavoro; l'art. 16 TFUE, in materia di protezione dei dati personali. La disposizione in materia di salute e sicurezza trova la propria base giuridica nell'art. 153; quella sulla sorveglianza umana anche nell'art. 16 TFUE. In altre parole, all'interno della direttiva piatta- forme, è il viatico offerto dalla prospettiva della protezione dei dati personali a consentire l'estensione della garanzia di contrasto alle discriminazioni algoritmiche a tutti i lavoratori. Con riguardo alla possibilità di estendere la protezione accordata dal diritto antidiscriminatorio ai lavoratori autonomi sulla base di una interpretazione sistematica delle fonti Ue, cfr. S. BORELLI-M. RANIERI, *La discriminazione nel lavoro autonomo. Riflessioni a partire dall'algoritmo Frank*, in *LLI*, 2021, n. 7(1), p. 21 ss.

incide piuttosto sull'eliminazione dei fatti costitutivi della fattispecie oggettiva, ovvero sulla garanzia di sussistenza dei fatti impeditivi. In tal senso, la valutazione di impatto periodica dovrebbe essere volta a eliminare eventuali differenze di trattamento, ovvero a garantire che le stesse o non si ricolleghino a fattori di rischio o siano scriminate dal fatto che le caratteristiche da cui derivano siano essenziali per lo svolgimento del lavoro ovvero siano appropriate e necessarie per raggiungere un legittimo obiettivo. Proprio questa seconda garanzia risulta particolarmente complessa nel caso di *machine learning*, se si considera che il funzionamento dell'apprendimento automatico si basa sull'individuazione di correlazioni e su calchi di *pattern* e non assicura, invece, quei nessi di consequenzialità logica necessari per applicare le citate scriminanti alle variabili utilizzate nel processo decisionale.

La medesima problematica incide, peraltro, sulla possibilità del datore di dimostrare la conformità del processo ai divieti di indagini su «fatti non rilevanti ai fini della valutazione dell'attitudine professionale del lavoratore», ovvero su «dati personali dei lavoratori che non siano strettamente attinenti alle loro attitudini professionali e al loro inserimento lavorativo», stabiliti rispettivamente dall'art. 8 St. Lav. e, per le agenzie per il lavoro, dall'art. 10, d.lgs. n. 276/2003[55].

Se è vero che il tema del contrasto alle discriminazioni sottolinea l'importanza – e quindi anche l'assenza – a livello normativo di un obbligo generale datoriale di valutazione di impatto dei processi automatizzati, le considerazioni svolte invitano a riflettere su quanto possa essere comunque rilevante, ai fini dell'adempimento degli obblighi datoriali e dell'effettività delle tutele da garantire, l'adozione su base volontaria di tale tecnica di controllo del rischio. Ossia, in altre parole, l'adempimento volontario dell'obbligo di valutazione di impatto sui diritti fondamentali di cui all'art 27 *AI Act*, al di fuori del suo ambito di applicazione[56].

È una prospettiva a cui invita lo stesso Regolamento, nel momento in cui valorizza i codici di condotta, quali strumenti comprensivi di un connesso meccanismo di *governance*, volti a promuovere l'applicazione volontaria di alcuni o tutti i requisiti obbligatori applicabili ai sistemi di IA ad alto rischio. Tali strumenti sono tanto più rilevanti nell'ottica lavoristica se si considerano, da un lato, l'importanza che assumono al loro interno i processi di alfabetizzazione, l'attenzione per le persone vulnerabili e la partecipazione dei portatori di interesse; dall'altro, la possibilità, prevista dall'art. 95, che tali codici siano elaborati dalle organizzazioni datoriali (in quanto rappresentative di *deployer*), con la partecipazione dei sindacati (rappresentativi dei portatori di interessi)[57].

[55] Cfr. TULLINI, *Divieto di indagini sulle opinioni e trattamenti discriminatori*, in *Il nuovo mercato del lavoro*, coordinato da M. PEDRAZZOLI, Zanichelli, Bologna, 2004, p. 145 ss. Dette disposizioni sono espressamente richiamate dall'art. 113, d.lgs. n. 196/2003, rubricato «Raccolta di dati e pertinenza», e rientrano tra le «norme più specifiche» per l'ambito dei rapporti di lavoro, rilevanti ai fini dell'art. 88 GDPR.

[56] Cfr. in tal senso anche A. MANTELERO, *The Fundamental Rights Impact Assessment (FRIA)*, cit.

[57] Cfr. S. CIUCCIOVINO, *op. cit.*

Come evidenzia la dottrina, la valorizzazione delle parti sociali nella configurazione dei sistemi organizzativi e gestionali, sostenuta dallo strumento dei codici, trova solido riscontro nella riflessione sul ruolo della "dimensione collettiva" nel controllo sull'effettività dei modelli di organizzazione e gestione di cui all'art. 30, d.lgs. n. 81/2008[58]. Il discorso riporta l'attenzione sugli esempi offerti dal quadro prevenzionistico, per evidenziare l'importanza, al suo interno, di strumenti che seppur facoltativi rappresentano uno sviluppo e un rafforzamento dell'approccio basato sul rischio, nonché «il metodo più idoneo per rispettare nel modo più adeguato» i precetti stabiliti dalla legge[59].

2. Le tecniche di giuridificazione del rischio nell'*AI Act* e i problemi derivanti dall'incerta delimitazione delle fattispecie normative

Nel corso della trattazione, si sono in più occasioni richiamate l'importanza e la funzionalità dell'*AI Act* nella costruzione del sistema integrato di fonti, a partire dall'utilizzo e dalla declinazione al suo interno dell'approccio basato sul rischio. Per completare il ragionamento, si ritiene significativo analizzare come tale approccio – e il connesso obiettivo di «introdurre un insieme proporzionato ed efficace di regole vincolanti» (cons. n. 26)[60] – trovi implementazione, all'interno del Regolamento, sul piano delle distinzioni per categorie di rischio (inaccettabile, alto e non-alto), posto che da tali distinzioni dipende la delimitazione del campo di applicazione delle garanzie esaminate.

In particolare, il Regolamento utilizza come tecnica di giuridificazione del rischio - a cui far corrispondere previsioni di divieto ovvero l'obbligo di rispettare il principale *corpus* di prescrizioni stabilite dal Capo III - l'individuazione di insiemi di pratiche d'uso e ipotesi di deroga tipizzate.

L'adattabilità del sistema è garantita dalla possibilità della Commissione di aggiornare e modificare, con atti delegati, la lista di pratiche d'uso stabilita dall'allegato III ai fini della classificazione ad alto rischio, laddove ricorrano le condizioni previste dall'art. 7.

---

[58] L. ZOPPOLI, *Il controllo collettivo sull'efficace attuazione del modello organizzativo diretto ad assicurare la sicurezza nei luoghi di lavoro*, in *WP Olympus*, 2012, n. 18; S. CIUCCIOVINO, *op. cit.*

[59] P. PASCUCCI, *Dopo il d.lgs. 81/2008: salute e sicurezza in un decennio di riforme del diritto del lavoro*, in P. PASCUCCI (a cura di), *Salute e sicurezza sul lavoro. Tutele universali e nuovi strumenti a 10 anni dal d.lgs. n. 81/2008*, Franco Angeli, Milano, p. 21. Sul punto cfr. M. GIOVANNONE, *op. cit.*; P. PASCUCCI-L. ANGELINI-C. LAZZARI, *I "sistemi" di vigilanza e di controllo nel diritto della salute e sicurezza sul lavoro*, in *Lavoro e diritto*, 2015, pp. 621 ss. Come evidenzia G. NATULLO, «il vero nodo nei percorsi di incremento della tutela della salute dei lavoratori sui luoghi di lavoro non è tanto nella astratta previsione normativa delle tutele stesse […] quanto nella garanzia di effettività di quelle tutele nei luoghi di lavoro. Maggiore effettività cui certamente sono funzionali la previsione di obbligatori e articolati percorsi procedurali […] (per tutti: la Valutazione dei rischi […]), così come la rilevanza prioritaria riconosciuta, anche a quei fini, a modelli organizzativi e gestionali "virtuosi"» (*Il quadro normativo dal Codice civile al Codice della sicurezza sul lavoro. Dalla Massima sicurezza (astrattamente) possibile alla Massima sicurezza ragionevolmente (concretamente) applicata?*, in *WP Olympus*, 2014, n. 39, p. 11).

[60] Per un'analisi dell'approccio basato sul rischio nell'*AI Act*, cfr. P. LOI, *Il rischio proporzionato nella proposta di regolamento sull'IA e i suoi effetti nel rapporto di lavoro*, in questa rivista, 2023, n. 4, pp. 239 ss.; M. BARBERA, *"La nave deve navigare". Rischio e responsabilità al tempo dell'impresa digitale*, in *LLI*, 2023, n. 9(2), p. 1 ss.

Si evidenziano, tuttavia, allo stato attuale, le criticità di un modello di giuridificazione del rischio che, da un lato, costruisce le categorie di rischio sulla base della tipizzazione di specifiche fattispecie, elemento affatto secondario nella logica dell'*AI Act* posta la centralità che assume al suo interno la finalità d'uso di un sistema ai fini dell'applicazione del quadro di regole (v. *infra*); dall'altro, tipizza le fattispecie e le ipotesi di deroga, affidandosi a enunciati normativi troppo generici o di dubbia interpretazione[61].

Esempi emblematici, nella prospettiva lavoristica, per quanto attiene alla categoria del rischio inaccettabile (art. 5), sono il divieto di sistemi per la valutazione o la classificazione delle persone fisiche o di gruppi di persone sulla base del loro comportamento sociale o di caratteristiche personali o della personalità, in cui il punteggio sociale (*social score*) così ottenuto comporti un trattamento pregiudizievole o sfavorevole (in un diverso contesto sociale ovvero ingiustificato o sproporzionato[62]); e il divieto di riconoscimento e inferenza delle emozioni sul posto di lavoro, salvo per motivi medici o di sicurezza[63].

Centrale sarà sul punto l'elaborazione da parte della Commissione di orientamenti sull'attuazione pratica dei divieti ai sensi dell'art. 96, par. 1, lett. b, attesi entro il 2 febbraio 2025, ossia entro la data di entrata in applicazione dell'art. 5. Si può peraltro segnalare, al riguardo, che la pubblicazione non è affatto prossima, visto che solo il 13 novembre 2024 la Commissione ha lanciato una consultazione agli stakeholder, indicando l'11 dicembre come termine di consegna dei feedback, al fine di elaborare successivamente le proprie linee guida.

Per quanto riguarda i sistemi di *social scoring*, la mancanza di una definizione normativa di "comportamento sociale" e di "punteggio sociale" interroga sull'ambito di applicazione della fattispecie. Ci si può chiedere, ad esempio, se rientri nel raggio d'azione del divieto un sistema algoritmico come quello adottato dalla *Caisse nationale des allocations familiales* (CNAF) francese per organizzare controlli mirati sui beneficiari delle prestazioni (come il reddito di solidarietà, l'assegno per disabili, l'indennità di attività a sostegno di lavoratori autonomi e subordinati ovvero di disoccupati). Al fine di individuare i fascicoli "sospetti", ossia quelli in cui è probabile siano riscontrati pagamenti indebiti, il sistema opera un processo di *data mining*, quindi di analisi automatica di *big data*, e sulla base di una molteplicità di variabili attribuisce mensilmente un punteggio di rischio: più alto è lo *score*, maggiore è la probabilità che l'interessato sia sottoposto a controllo. Dal ricorso al Consiglio di Stato presentato a ottobre 2024 da 15 organizzazioni della società civile contro l'utilizzo di questo processo decisionale automatizzato[64], si evince come i dati personali

[61] Si consenta di rinviare sul punto anche A. MANTELERO-M. PERUZZI, *L'AI Act e la gestione del rischio nel sistema integrato delle fonti*, in *RGL*, 2024, n. 4, in corso di pubblicazione.

[62] A norma dell'art. 5, par. 1, lett. c), tale trattamento, per rilevare ai fini dell'integrazione della fattispecie, deve o avvenire «in contesti sociali che non sono collegati ai contesti in cui i dati sono stati originariamente generati o raccolti» o essere «ingiustificato o sproporzionato rispetto al […] comportamento sociale o alla sua gravità».

[63] Cfr. al riguardo anche M. BIASI, *Problema e sistema nella regolazione lavoristica dell'intelligenza artificiale: note preliminari*, in questa rivista, 2024, in corso di pubblicazione.

[64] V. *L'algorithme de notation de la Cnaf attaqué devant le Conseil d'État par 15 organisations*, su www.gisti.org

trattati riguardino sia i beneficiari delle prestazioni sia i loro familiari (si tratta di oltre 32 milioni di persone) e le variabili ponderate dall'algoritmo spazino dalla situazione familiare, occupazionale e finanziaria del soggetto alle caratteristiche sociali ed economiche del suo comune di residenza, dal numero di variazioni dello stato di famiglia al numero di variazioni del luogo di residenza. Al di là dei profili di rilievo che questo tipo di pratica può presentare per violazione del GDPR e della normativa antidiscriminatoria (evidenziati nel citato ricorso), ci si può d'altra parte domandare se le modalità con cui viene attribuito il punteggio di rischio non configurino un social scoring, se i dati trattati non traccino il comportamento sociale del soggetto e se quindi il processo non sia a monte da vietare, in quanto afferente alla categoria del rischio inaccettabile di cui all'art. 5 dell'*AI Act*, applicabile già a partire da febbraio 2025.

Per quanto attiene alla seconda ipotesi di divieto menzionata, ossia quella relativa all'utilizzo di sistemi di riconoscimento delle emozioni sul luogo di lavoro (art. 5, par. 1, lett. f), si consideri che tali sistemi sono definiti, all'art. 3, come sistemi finalizzati «all'identificazione o all'inferenza di emozioni o intenzioni di persone fisiche sulla base dei loro dati biometrici», questi ultimi a loro volta qualificati, sempre all'art. 3, come «dati personali ottenuti da un trattamento tecnico specifico relativi alle caratteristiche fisiche, fisiologiche o comportamentali di una persona fisica, quali le immagini facciali o i dati dattiloscopici» (la nozione di «dati biometrici» dovrebbe essere comunque interpretata alla luce di quella prevista nel GDPR, secondo il cons. n. 14). Il punto cruciale attiene all'ambito di applicazione della deroga al divieto, prevista per l'uso giustificato da motivi medici o di sicurezza. Su tale profilo, infatti, si riscontra già un aperto contrasto interpretativo tra quanto pubblicato dalla Commissione europea sul sito istituzionale nella sezione "Domande e risposte", aggiornata ad agosto 2024, e il considerando n. 18. Anche la formulazione di quest'ultimo ha visto, invero, una significativa evoluzione dal testo di compromesso interistituzionale adottato a dicembre 2023 al testo approvato dal Pe a fine aprile 2024 (e quindi poi dal Consiglio), un'evoluzione che risolvendo una contraddittorietà interna[65] ha lasciato invece sospeso il contrasto con quanto affermato dalla Commissione. Segnatamente, nel considerando si afferma che «la nozione di "sistema di riconoscimento delle emozioni" [...] non comprende stati fisici, quali dolore o affaticamento, compresi, ad esempio, ai sistemi utilizzati per rilevare lo stato di affaticamento dei piloti o dei conducenti

[65] Il testo di compromesso interistituzionale, al cons. 8a, prevedeva il seguente testo, che si riporta in inglese per consentire di meglio comprendere la criticità di formulazione: «*The notion of emotion recognition system [...] It does not include physical states, such as pain or fatigue. It refers for example to systems used in detecting the state of fatigue of professional pilots or drivers for the purpose of preventing accidents*». Come si può notare, nel primo te- sto adottato, si affermava che la nozione di sistema di riconoscimento delle emozioni, da un lato, non include stati fisici come il dolore e la fatica, dall'altro, si riferisce a titolo di esempio ai sistemi di rilevamento della fatica di piloti e autisti. Successivamente, l'ultima frase citata è stata modificata nel modo seguente: «*It does not include physical states, such as pain or fatigue; this refers for example to systems used in detecting the state of fatigue...*». La frase è stata ulteriormente rielaborata nel testo finale: «*It does not include physical states, such as pain or fatigue, including, for example, systems used in detecting the state of fatigue of professional pilots or drivers for the purpose of preventing accidents*».

professionisti al fine di prevenire gli incidenti»; nella sezione "Domande e risposte", la Commissione individua come esempio di deroga al divieto, giustificata da motivi di sicurezza, proprio «il monitoraggio dei livelli di stanchezza di un pilota». Il contrasto è tutt'altro che irrilevante, posto che la Commissione, come detto, ai sensi dell'art. 96 dell'*AI Act*, elabora gli orientamenti per l'attuazione pratica dei divieti di cui all'art. 5: è importante quindi verificare se la posizione sarà confermata all'interno di detti orientamenti. La delimitazione del perimetro della nozione di "sistema di riconoscimento delle emozioni" e, di conseguenza, delle pratiche che possono rientrare nella deroga al divieto è rilevante anche per l'individuazione del regime a cui tali pratiche, pur ammesse, sarebbero soggette. I sistemi di riconoscimento delle emozioni sono individuati, infatti, dall'allegato III come una autonoma e specifica area critica d'uso (al punto 1) ai fini della classificazione nella categoria ad alto rischio e devono essere sottoposti a uno stringente processo di valutazione di conformità: ai sensi dell'art. 43, par. 1, il mero controllo interno è consentito, infatti, solo a condizione che siano state previamente introdotte, nonché applicate dal fornitore, norme armonizzate (art. 40) o specifiche comuni (art. 41) volte a sostenere e garantire la *compliance* con una declinazione ad hoc dei requisiti richiesti[66]. Seguendo la linea interpretativa (per ora) adottata dalla Commissione, rientrerebbero nell'ambito della deroga al divieto di riconoscimento delle emozioni sul luogo di lavoro (e sarebbero sottoposti pertanto al regime previsto per tali pratiche) i sistemi che per motivi di sicurezza e, in particolare, per rilevare un eventuale stato di affaticamento, stress o distrazione catturino attraverso sensori e trattino informazioni sulla posizione e diametro della pupilla, vettore e punto dello sguardo, postura, andatura e posizione della testa, frequenza cardiaca e respiratoria, sudorazione, temperatura corporea, tono della voce, espressione del viso, linguaggio verbale o scritto, battitura sulla tastiera. Diversamente, seguendo la linea adottata dal cons. n. 18, tali sistemi, essendo finalizzati a rilevare uno stato fisico, resterebbero esclusi dalla nozione di sistema di riconoscimento delle emozioni. Una loro classificazione come ad alto rischio potrebbe pertanto avvenire solo per afferenza ad un'altra area critica d'uso individuata dall'allegato III, quella specificamente dedicata all'ambiente di lavoro (punto n. 4). Tale area, tuttavia, prevede una lista di finalità d'uso che non include, almeno esplicitamente, quella di tutela della sicurezza. In essa si fa riferimento a sistemi destinati a essere utilizzati «per l'assunzione o la selezione di persone fisiche» ovvero «per adottare decisioni riguardanti le condizioni dei rapporti di lavoro, la promozione o cessazione dei rapporti contrattuali di lavoro, per assegnare compiti sulla base del comportamento individuale o dei tratti e delle caratteristiche personali o per monitorare e valutare le prestazioni e il comportamento delle persone».
La questione è tutt'altro che secondaria nel contesto dell'*AI Act*, posto che è proprio la finalità d'uso a

[66] Le prime sono adottate dalle organizzazioni europee di normazione su richiesta della Commissione, le seconde sono specifiche per detti requisiti stabilite da atti di esecuzione della Commissione, adottati in assenza o carenza delle citate norme armonizzate.

caratterizzare un sistema di IA e la valutazione dei rischi che comporta; da essa dipendono la classificazione nelle categorie di rischio e l'applicazione delle corrispondenti garanzie (v. art. 8, par. 1)[67]. Non è un caso che la finalità, come precisato già dall'art. 3, debba essere comunicata e dettagliata nelle istruzioni d'uso e, se cambiata, comporti una «modifica sostanziale» del sistema di IA, con passaggio in capo a chi l'ha apportata della posizione del fornitore, con relativi obblighi, ai sensi dell'art. 25. Significativamente, sempre la finalità d'uso prevista dovrebbe rientrare nell'informativa che il *deployer* deve garantire nei confronti delle persone fisiche soggette all'uso del sistema (nonché dei loro rappresentanti, se il *deployer* è datore di lavoro e le persone sono lavoratori; v. art. 26, parr. 7 e 11; cons. n. 93)[68].

Le criticità derivanti dalla formulazione dell'allegato III, peraltro modificabile dalla stessa Commissione ai sensi dell'art. 7, potrebbero essere superate ragionando sulla stretta connessione che si dovesse instaurare tra la finalità di sicurezza, per cui è adottato il sistema, e alcune funzioni indicate nel punto 4, come l'adozione di decisioni sulle condizioni dei rapporti di lavoro, l'assegnazione dei compiti o il monitoraggio del comportamento[69].

Rimane da segnalare il diverso regime a cui i sistemi elencati al punto 4 dell'allegato III sono sottoposti, per quanto riguarda la valutazione di conformità: se, infatti, per quelli di riconoscimento delle emozioni, come detto, il controllo interno è ammesso solo a determinate condizioni, per i menzionati sistemi di gestione dei lavoratori il processo di valutazione – nonostante la forte critica avanzata dal fronte sindacale – è sempre eseguibile da parte del fornitore sulla base di un mero *self-assessment* (art. 43, par. 2).

Si può, altresì, rilevare come la linea tracciata dal cons. n. 18 non solo trovi scarso riscontro nella pratica e nella letteratura relative ai sistemi algoritmici di riconoscimento delle emozioni, ove si evidenzia il ruolo chiave di tali sistemi nella rilevazione di fatica, sonnolenza, dolore, depressione e *burn-out* e si riporta, ancora una volta, tra gli esempi, il monitoraggio dello stato di affaticamento della persona alla guida, attraverso l'analisi delle espressioni facciali, dei movimenti degli occhi e di vari segnali fisici e fisiologici[70]. Tale linea non consente nemmeno di comprendere quali siano le ipotesi rilevanti ai fini dell'applicazione

[67] Come segnalato dal cons. n. 52, «per quanto riguarda i sistemi di IA indipendenti, ossia [...] diversi da quelli che sono componenti di sicurezza o che sono essi stessi prodotti» (e in quanto tali soggetti anche alla relativa normativa Ue), la valutazione del rischio, da cui dipende la classificazione, è effettuata «alla luce della loro finalità prevista».

[68] Cfr. in tal senso anche L. TEBANO, *Intelligenza Artificiale e datore di lavoro: scenari e regole*, in *DLM*, 2024, in corso di pubblicazione.

[69] Cfr. in tal senso S. MARASSI, *Intelligenza artificiale e sicurezza sul lavoro*, in M. BIASI (a cura di), *Diritto del lavoro e intelligenza artificiale*, Giuffré, Milano, 2024 p. 207 ss. Rimane ferma la possibilità di classificare il sistema come ad alto rischio anche in base al- l'art. 6, par. 1, qualora sia componente di sicurezza di prodotti o esso stesso prodotto e in forza di una delle fonti Ue elencate nell'allegato I – tra cui rientra non solo il nuovo Regolamento macchine 2023/1230 ma anche il Regolamento 2016/425 sui dispositivi di protezione individuale – sia sottoposto a una procedura di valutazione di conformità da parte di organismi terzi.

[70] Cfr. S.K. KHARE-V. BLANES-VIDAL-E.S. NADIMI-U. RAJENDRA ACHRAYA, *Emotion recognition and artificial intelligence: A systemic review (2014-2023) and research recommendations*, in *Information Fusion*, 102, 2024, 1 ss.

della deroga, non essendo offerto a tal proposito alcun supporto dal successivo cons. n. 44, che individua come unico esempio «i sistemi destinati all'uso terapeutico».

L'utilizzo della tipizzazione ai fini della giuridificazione del rischio presenta le stesse problematiche, in termini di costruzione delle fattispecie normative, con riguardo all'applicazione dei criteri di deroga alla classificazione ad alto rischio previsti dall'art. 6, par. 3, in particolare nella definizione delle condizioni che escludono il rischio significativo di danno e consentono di non considerare "ad alto rischio" le fattispecie tipizzate nell'allegato III, tra cui quelle di cui al punto 4, relative all'occupazione, gestione dei lavoratori e accesso al lavoro autonomo[71].

Si consenta brevemente di ricordare come a norma dell'art. 6, par. 3, il rischio significativo di danno sia ritenuto escluso laddove il sistema non influenzi materialmente il risultato del processo decisionale, ovvero sempre presente qualora il sistema effettui profilazione di persone fisiche.

Rispetto all'ipotesi di esclusione del rischio citata, pare opportuno rilevare come il grado di impatto sul processo decisionale a cui la disposizione fa riferimento non coincida necessariamente con l'integrale automatizzazione dello stesso. Come spiega il cons. n. 53, «un sistema di IA che non influenza materialmente l'esito del processo decisionale dovrebbe essere inteso come un sistema di IA che non ha un impatto sulla sostanza, e quindi sull'esito, del processo decisionale, sia esso umano o automatizzato». Da ciò si può trarre che il rischio sussista anche laddove il sistema abbia un impatto sull'esito di un processo decisionale umano, ossia laddove sia impiegato, per utilizzare le parole della direttiva piattaforme, non per prendere ma (solo) per sostenere una decisione. A conferma si pone l'obbligo di informazione del *deployer*, previsto dall'art. 26, par. 11, in caso di sistemi ad alto rischio che adottano decisioni o *assistono* nell'adozione di decisioni.

Sempre il par. 3 tipizza, d'altra parte, le condizioni che escludono il rischio, foriere di non pochi dubbi interpretativi, soprattutto se raffrontate con alcune aree critiche d'uso elencate dall'allegato III.

Il rischio significativo di danno sarebbe, ad esempio, da escludere, qualora il sistema fosse solo volto a rilevare lo scostamento di una valutazione umana già completata rispetto a uno schema prestabilito, potendo procedere a una sua sostituzione o determinando un condizionamento della stessa solo con adeguata revisione umana. L'indicazione interroga sui termini con cui la presenza della revisione umana in un processo decisionale automatizzato sia in grado di escludere l'alto rischio (quando cioè rilevi e si possa ritenere di misura adeguata).

Parimenti il rischio è escluso se il sistema di IA esegue un compito procedurale limitato (ad es. classificazione di documenti per categorie), migliora il risultato di un'attività umana precedentemente completata (ad es. allineamento del testo di un documento già redatto a un determinato stile), ovvero

[71] Cfr. al riguardo anche A. ALAIMO, *op. cit.*

esegue un compito solo preparatorio per una valutazione pertinente ai casi d'uso di cui all'allegato III, quindi ad esempio ai fini della selezione o gestione del personale. Il cons. n. 53 segnala, tra i possibili compiti preparatori, «soluzioni intelligenti per la gestione dei fascicoli, che comprendono varie funzioni quali […] il collegamento dei dati ad altre fonti di dati». La delimitazione della fattispecie andrebbe però meglio chiarita, perché il generico enunciato normativo non consente di capire se o quando, ad esempio, uno *screening* di candidature effettuato da un sistema di IA, indicato tra le finalità d'uso critiche dell'allegato III (sistemi utilizzati «per filtrare le candidature e valutare i candidati»), possa integrare la condizione di deroga alla classificazione ad alto rischio, in quanto compito prodromico a restringere la rosa di candidati da sottoporre a una valutazione umana. Il tema è tanto più rilevante considerati i rischi di *proxy discrimination* che tali processi di pre-selezione possono implicare.

Anche in questo caso, bisognerà attendere, a norma dell'art. 6, par. 5, gli orientamenti della Commissione, questi da fornire entro il 2 febbraio 2026, previa consultazione del Consiglio europeo per l'IA.

Ad ogni modo, quantomeno nell'esempio riportato, pare difficile escludere che il processo, riducendo l'ambito della scelta, non influenzi materialmente la decisione umana finale. Soprattutto, a risolvere a monte molte questioni interpretative poste dalla costruzione delle ipotesi di deroga dovrebbe essere l'applicabilità dell'ipotesi di conferma del rischio sempre stabilita dal par. 3, ossia la profilazione di persone fisiche. Per la definizione, l'*AI Act* rinvia a quanto stabilito nel GDPR, che qualifica la fattispecie come un trattamento automatizzato finalizzato a valutare, ai fini di analisi e/o previsioni, aspetti e caratteristiche personali individuali, come il rendimento, l'ubicazione, l'affidabilità, il comportamento. Si può evidenziare come sia molto probabile che tale condizione ricorra in caso di utilizzo di processi automatizzati di monitoraggio o decisionali sul lavoro, con ciò potendosi ritenere tendenzialmente da escludere una classificazione come "a basso rischio" dei sistemi destinati a essere utilizzati nell'esercizio di prerogative datoriali nei confronti dei lavoratori.