# APPLICAZIONE DELL'INTELLIGENZA ARTIFICIALE NELLA PUBBLICA AMMINISTRAZIONE

RISULTATI DEL GDL PER LA IA DEL 2021

a cura di: DIGITAL&LAW

# Sommario

# *Introduzione*

*In questi anni stiamo assistendo ad un'autentica opera di trasformazione, il cosiddetto processo di "digitalizzazione" della Pubblica Amministrazione. Si tratta di un percorso di rinnovamento in primis degli strumenti amministrativi, che deve fare i conti non solo con la reingegnerizzazione documentale, ma anche con la sempre più pervasiva adozione dell'Intelligenza Artificiale. Alcune soluzioni, quali la profilazione degli utenti, l'elaborazione di processi predittivi in determinati ambiti, il riconoscimento, l'indicizzazione e la classificazione delle informazioni, fanno parte della prassi amministrativa da diverso tempo.*

*Accanto agli strumenti, si collocano poi le competenze degli attori coinvolti in prima persona. Secondo l'ultimo censimento delle istituzioni da parte dell'Istat (con dati aggiornati al 31 dicembre 2018) in Italia lavorano per il settore pubblico 3.457.498 dipendenti. Non a caso, il Piano Nazionale di Ripresa e Resilienza individua nelle persone, prima ancora che nelle tecnologie, il motore del cambiamento e dell'innovazione nella Pubblica amministrazione.*

*Sulla scia delle iniziative intraprese per l'aggiornamento e la formazione del personale pubblico, nel settembre del 2021 ANORC ha intrapreso la conduzione di un GDL dedicato all'applicazione dell'Intelligenza Artificiale nel settore pubblico. Abbiamo coinvolto specialisti con campi di interesse eterogeni e approcci diversificati per dar vita ad un confronto sulle possibili applicazioni strategiche dell'IA a livello nazionale, passando attraverso l'analisi dei diversi fattori da tenere in considerazione e senza tralasciare il confronto con altri scenari di matrice europea.*

*Abbiamo chiesto loro di affrontare il tema principale attraverso la propria formazione, esperienza e sensibilità, per mettere gli argomenti sul tavolo. Il presente lavoro intende favorire la circolazione del dibattito sul tema, uscendo dai confini del gruppo di lavoro per rendere accessibili questi diversi punti di vista sulla complessità dello sforzo di rinnovamento che sta investendo la pubblica amministrazione.*

*Buona lettura da ANORC.*

**Componenti del GDL IA 2021**: *Prof. Michele Iaselli (Coordinamento); Avv. Anna Capoluongo; dott. Luigi Meroni; ing. Andrea Piccoli; dott.ssa Valentina Sapuppo; dott.ssa Flora Tozzi*

*Lecce, marzo 2022*

*Grafica:* **Marcello Moscara**

*Edizione e impaginazione a cura di:* **Digital&Law**

# *Legge sull'intelligenza artificiale: analisi delle tipologie di rischio*

*di Flora Tozzi*

## *Premessa*

*Nell'ambito della strategia europea per l'Intelligenza Artificiale, la Commissione europea ha pubblicato il 21 aprile 2021, la proposta di Regolamento sull'approccio europeo all'Intelligenza Artificiale [COM(2021) 206 final], che propone il primo quadro giuridico europeo sull'IA: Artificial Intelligence Act. La proposta contiene delle norme orizzontali per lo sviluppo, la commercializzazione e l'uso dei servizi e sistemi basati sull'IA.*

*La Commissione europea ha scelto di* **definire i sistemi di intelligenza artificiale e non l'IA** *definendo "un software sviluppato con una o più delle tecniche e degli approcci elencati nell'allegato I e che può, per un determinato insieme di obiettivi definiti dall'uomo, generare risultati quali contenuti, previsioni, raccomandazioni o decisioni che influenzano gli ambienti con cui interagiscono" (cfr. art. 3,della proposta di Regolamento sull'IA) ,includendo tecniche e approcci informatici di IA:*

*a) Approcci di apprendimento automatico, compresi l'apprendimento supervisionato, l'apprendimento non supervisionato e l'apprendimento per rinforzo, con utilizzo di un'ampia gamma di metodi, tra cui l'apprendimento profondo (deep learning);*

*b) approcci basati sulla logica e approcci basati sulla conoscenza, compresi la rappresentazione della conoscenza, la programmazione induttiva (logica), le basi di conoscenze, i motori inferenziali e deduttivi, il ragionamento (simbolico) e i sistemi esperti;*

*c) approcci statistici, stima bayesiana, metodi di ricerca e ottimizzazione", (cfr. Allegato I, proposta di Regolamento sull'IA).*

*Questo approccio non deterministico può creare incertezze da parte degli sviluppatori/fornitori di sistemi di IA. Anche se dal considerando 6 si percepisce che la Commissione abbia voluto dare una definizione più particolareggiata: "la definizione di sistema di IA dovrebbe essere definita in maniera chiara al fine di garantire la certezza del diritto, prevedendo nel contempo la flessibilità necessaria per agevolare i futuri sviluppi tecnologici. La definizione dovrebbe essere basata sulle principali caratteristiche funzionali del software, in particolare sulla capacità, per una determinata serie di obiettivi predisposti dall'uomo, di generare output quali:*

- *contenuti*
- *previsioni*
- *raccomandazioni o decisioni, che influenzano l'ambiente con cui il sistema interagisce, "sia nella dimensione fisica che digitale".*

*Invece, esaminando l'allegato I la circostanziata asserzione della definizione del considerando 6 fa riferimento sia ad* **un approccio sull' apprendimento supervisionato, non supervisionato** *come machine learning,* **sull'apprendimento profondo** *come il deep learning e sia ad* **un approccio basato sulla logica e sulla conoscenza**, *tra cui: rappresentazione della conoscenza, programmazione logica/induttiva, ragionamento simbolico e sistemi esperti, approcci statistici, stima bayesiana per finire con metodi di ricerca e ottimizzazione.*

*Da quest'ultima definizione, alquanto ampia, si desume, che la Commissione europea abbia voluto ricomprendere i sistemi di intelligenza artificiale in forma:*

- **debole***: sistemi che possono agire e pensare come se avessero capacità di pensiero, ma senza averla. Si tratta di IA che ha capacità di calcolo e di memoria, lì dove nessuna persona è capace di fronteggiare delle operazioni*[1]*complesse.*
- **forte***: sistemi che possono pensare come una vera e propria mente umana, quindi hanno la capacità cognitiva e la conoscenza delle proprie capacità e limiti.*

*La prima forma viene impiegata per* **applicazioni piuttosto innocue** *come giochi e assistenti vocali, mentre la seconda per* **applicazioni critiche** *come la sicurezza dei sistemi di assistenza nella guida, i sistemi di rilevamento delle intrusioni e nella diagnosi delle malattie.*

*Nel momento in cui si utilizzano per fini rivolti al benessere dell'uomo devono dimostrare dei requisiti abbastanza forti, tali da instaurare il giusto ambiente di fiducia per lo sviluppo e un impiego efficace, basandosi sul principio di qualità e sicurezza*[1]*.*

*L'obbiettivo principale della Commissione europea è di garantire che tutti i sistemi di IA utilizzati siano sicuri, trasparenti, etici, imparziali e sotto il controllo umano*[2]*, affinché ad essi venga associato il concetto di affidabilità.*

***Il concetto di affidabilità è dato dal binomio di responsabilità e obblighi*** *che sono stati trasferiti dagli esseri umani all'IA in materia di sicurezza e protezione. È indispensabile che, oltre ad essere affidabili, tali sistemi siano sufficientemente sicuri da dimostrarsi* **resilienti sia agli attacchi**, *che ai tentativi di manipolazione di dati o algoritmi, avendo* **un piano di emergenza** *in caso di problemi, allo scopo di valutare quali decisioni prendere e a quali risultati portano.*

*Altro elemento da prendere in considerazione nell'ambito dell'affidabilità è la* **conoscenza dei processi di progettazione**, *in quanto l'uomo presenta delle difficoltà nel capire perché si è comportato in un certo modo e perché ha fornito una data interpretazione. La sfida che deve essere affrontata è quella di cercare di comprendere meglio i meccanismi alla base del sistema IA, in particole quelli basati sulle reti neurali. Da processi di addestramento con reti neurali possono risultare parametri di rete*

[1] Cfr. *La carta etica europea sull'utilizzo dell'intelligenza artificiale nei sistemi giudiziari e negli ambiti connessi.* (CEPEJ(2018)14), pag.10.
[2] Cfr. considerando 6, proposta di Regolamento sull'IA.

*impostati su valori numerici difficilmente correlabili con i risultati; quindi, la rappresentazione interna dei dati potrebbero creare una classificazione degli input diversa dalle risposte in uscita.*

*Occorre quindi conoscere fino a che punto è possibile effettuare dei meccanismi di sorveglianza gestiti da processi di* **supervisione umana** *(human on the loop) o da un* **controllo umano** *(human in command).*

*Percorrendo tra le sfaccettature dell'affidabilità occorre che si includa anche l'elemento "***prestazioni***" che dipende fortemente dalla qualità e dalla quantità dei dati immessi o elaborati, che rappresentano la materia prima dei processi di addestramento degli algoritmi; nelle reti neurali i processi di progettazione non sono ben definiti e pertanto questi ultimi possiedono una* **vulnerabilità qualitativamente nuova**, *tanto da essere oggetto di attacchi da parte degli aggressori (in particolar modo nel campo della sanità e finanziaria).*

*Per evitare tale vulnerabilità occorre esaltare i requisiti della* **riservatezza, governance dei dati e integrità**, *indispensabili in tutte le fasi del ciclo di vita del sistema IA, nei processi e nei set di dati utilizzati, il tutto opportunamente testato e documentato in ogni fase, come ad esempio: la pianificazione, l'addestramento, i test e la diffusione.*

*In sintesi, i sistemi di IA devono essere implementati e realizzati su dei requisiti fondamentali, tanto da essere stati accolti anche dalla Commissione UE*[3]*, per poter essere considerati affidabili, sicuri e conformi alle norme e ai valori europei.*


*Figura 1- i sette requisiti di un IA affidabile, riportati nel documento "Le Linee guida etiche per l'intelligenza artificiale affidabile (AI) "della Commissione europea, 2019.*

[3] Cfr. Le Linee guida etiche per l'intelligenza artificiale affidabile (AI). <https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html>,2019.

## *Creare Fiducia nell'IA antropocentrica*

*L'UE continua a lavorare per creare un ambiente che stimoli gli investimenti finalizzati al "piano di sviluppo dell'IA", affinché si generi competitività della tecnologia nelle comunità di ricerca all'avanguardia nel mondo, negli imprenditori innovativi e nelle imprese start-up ad elevatissimo contenuto tecnologico. In questo senso l'UE ha inteso favorire lo sviluppo e l'utilizzo delle piattaforme che forniscono servizi verso imprese intelligenti.*

*La strategia messa in atto dall'UE deve considerare che vi è, anche,* **un coinvolgimento dell'uomo e/o del cittadino**, *ad esempio nelle decisioni amministrative, e solo esaltando l'elemento "fiducia", indispensabile per assicurare che l'IA non sia fine a sé stessa, può diventare uno strumento a servizio delle persone e dei cittadini con lo scopo di migliorarne il loro benessere. Visto che il beneficiario dei servizi è l'uomo è fondamentale proporre e suscitare azioni che conducono al* **principio antropocentrico**. *Di qui si aprono altre questioni, come l'etica che potrà derivare dall'uso di tale tecnologia; la neutralità dei dati, la trasparenza degli algoritmi e la responsabilità dei soggetti che vi fanno ricorso.*

***La fiducia è una condizione indispensabile*** *per assicurare un avvicinamento antropocentrico all'IA. Dove i valori, su cui si fonda l'Unione, sono rivolti al rispetto della dignità umana, della libertà, della democrazia e dell'uguaglianza; il rispetto dei valori lo si ottiene attraverso un solido quadro normativo destinato a diventare lo standard internazionale per un'IA antropocentrica.*

*A tale proposito la Commissione aveva già espresso la propria opinione nel comunicato intitolato "Creare fiducia nell'intelligenza artificiale antropocentrica", in cui: "I sistemi di IA dovrebbero inoltre contenere meccanismi di sicurezza fin dalla progettazione, per garantire che siano sicuri in modo verificabile in ogni fase, considerando soprattutto la sicurezza fisica e mentale di tutte le persone coinvolte. Ciò comprende anche la possibilità di ridurre al minimo e, ove possibile, rendere reversibili gli effetti involontari o gli errori del funzionamento del sistema. È opportuno prevedere processi in grado di chiarire e valutare i potenziali rischi associati all'uso dei sistemi di IA nei vari settori di applicazione."*

## *Un approccio basato sul rischio*

*Il legislatore, in questa proposta, ha voluto differenziare le tecnologie che a seconda delle circostanze, delle applicazioni e dell'utilizzo, possono comportare dei rischi per i diritti fondamentali e la sicurezza dell'uomo.*

*Ad esempio, lo strumento di filtraggio dello spam basato sulle statistiche bayesiane, non sarà soggetta ai nuovi requisiti al contrario delle reti bayesiane utilizzate per il*

*triage dei pazienti nel pronto soccorso, che presentano dei rischi non solo in ambito sanitario, ma anche sociale.*

*Queste differenziazioni fanno sì che la proposta di Regolamento sull'IA sia basata sul rischio e su un approccio proporzionato che garantisca* ***condizioni di parità e di protezione efficace dei diritti e delle libertà delle persone in tutta l'Unione.***


*Figura 2-Quadro giuridico affidabile dell'IA*

*Quindi le caratteristiche di affidabilità e di fiducia dei sistemi di IA sono riconosciute solo attraverso un quadro di sicurezza dei prodotti/servizi IA, il quale si erige su* ***quattro livelli di rischio.***

*L'approccio basato sul rischio, già noto nel GDPR, impone* ***oneri e gradi di tutele crescenti*** *all'aumentare del livello di rischio per la sicurezza, la salute e i diritti fondamentali dei soggetti, basandosi alle singole tipologie di IA.*

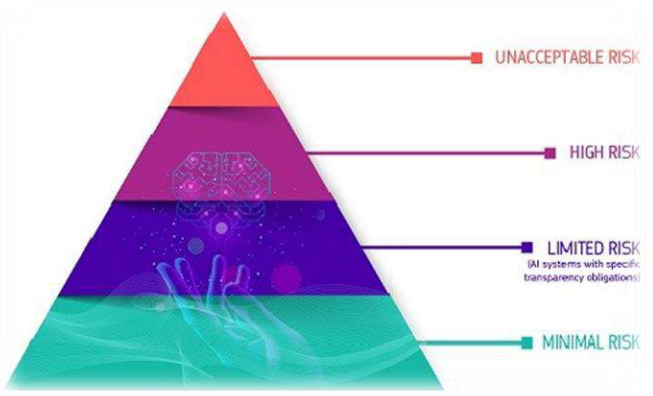



*Figura 3 - Il rischio dell'IA è rappresentabile in una piramide dove al vertice si trovano i sistemi a rischio inaccettabile, nel centro quelli a rischio elevato e limitato e alla base quelli a rischio minimo*

*L'approccio corretto verso un sistema di individuazione e attribuzione del corretto livello dei sistemi di IA è sicuramente la chiave del successo che porta anche ad affrontare problematiche legate all'etica, al sociale e all'ambiente.*

## *I tipi di rischi*

### Rischio inaccettabile

*Vengono individuati ed elencati della proposta di Regolamento sull'IA specifiche pratiche di IA (piuttosto che sistemi) che possono generare un rischio inaccettabile. Sono vietate le pratiche che presentano un potenziale manipolazione delle persone attraverso tecniche subliminali, senza che i destinatari ne siano consapevoli, oppure di sfruttamento delle vulnerabilità su specifiche categorie di persone, quali minori o disabili, al fine di* **distorcerne materialmente il comportamento** *in maniera tale da provocare a loro o a un altro interessato un danno psicologico o fisico.*

*Le pratiche di IA a rischio inaccettabile sono state individuate in quattro categorie, di cui:*

- *due riguardano sistemi cognitivi di manipolazione comportamentale di persone o specifici gruppi vulnerabili;*
- *due fanno riferimento a social scoring e sistemi di identificazione biometrica in tempo reale da remoto.*

*Nel primo gruppo fanno parte, ad esempio, quelli che attraverso il* **controllo delle emozioni** *possono portare alla manipolazione comportamentale di bambini, anziani o altri consumatori, sfruttandone le vulnerabilità cognitive e inducendoli a scelte commerciali indesiderate. Di fondo c'è qualcosa che preoccupa non solo lo scienziato ma lo giurista ed è il problema della scatola nera: algoritmi opachi e processi di difficile contestazione.*

*Mentre i sistemi di IA di social scoring raccolgono i dati correlati a un determinato ambito per categorizzare l'interessato in relazione a servizi riferibili in ambiti diversi rispetto al primo. Si pensi all'utilizzo dei dati di videosorveglianza raccolti su un treno per attribuire un punteggio sociale al cittadino a cui possa poi farsi riferimento per consentire o inibirne l'accesso a manifestazioni sportive. Dovrebbe essere vietato anche il social scoring basato sul modello cinese.*

*Anche se l'art. 5 della proposta di Regolamento sull'IA delinea le aree di applicazione ma c'è da dire che il diavolo è nei dettagli a causa di un ampio margine di interpretazione*[4]*. Invece, la definizione di casi d'uso specificamente vietati sarebbe molto utile (ad esempio utilizzare algoritmi che vietano la visualizzazione di contenuti*

[4] Ad esempio, anche il funzionamento di un motore di ricerca su una banca dati potenzialmente non rappresentativa potrebbe essere vietato dall'articolo 5, lettera a) , a causa del potenziale impatto dei risultati della ricerca.

*a ragazzi inferiori di 13 anni su piattaforme come Instagram), e dovrebbe essere esaminato, anche, se questi divieti siano esplicitamente vietati.*


*Figura 4-Elenco pratiche di intelligenza artificiale vietate*

## Rischio alto

*Oltre agli usi vietati, la Commissione europea definisce una serie di usi associati a un rischio elevato ma che restano consentiti in base a determinate condizioni. Il titolo III della proposta di Regolamento sull'IA definisce i sistemi di IA come "ad alto rischio" ma per essere più precisi definisce le "Regole di classificazione per i sistemi di IA ad alto rischio".*

*La classificazione dei sistemi di IA ad alto rischio raggruppa due categorie principali:*

- *sistemi di IA da utilizzare come componenti di sicurezza dei prodotti;*
- *sistemi di IA autonomi.*

*Il primo gruppo si riferisce ai sistemi di IA che sono fisicamente integrati nel prodotto (embedded), mentre il secondo funziona in modo indipendente senza essere integrato nel prodotto (non embedded).*

*Questa classificazione è importante, perché l'UE vuole applicare questa nuova normativa in questione di sistemi di IA ad alto rischio indipendentemente dalla loro ubicazione. In altre parole, il Regolamento sull'IA sarebbe applicabile ai fornitori e agli utenti stabiliti in un paese terzo come, ad esempio, Cina e Stati Uniti, a condizione che l'output prodotto da tali sistemi di IA sia utilizzato all'interno dell'UE.*

*Tra gli esempi di sistemi di IA ad alto rischio troviamo quelli associali ai componenti di sicurezza di un prodotto o di un sistema, l'art. 3 della proposta di Regolamento sull'IA definisce il componente di sicurezza di un prodotto o di un sistema come: "un*

*componente di un prodotto o di un sistema che svolge una funzione di sicurezza per tale prodotto o sistema il cui guasto o malfunzionamento mette in pericolo la salute e la sicurezza di persone o beni"; in questo quadro normativo non è bene chiara la funzione di sicurezza, cosa intende specificatamente la Commissione europea per applicazioni "sicure". Inoltre, non è specificato come la "sicurezza"[5] possa essere stabilita nei singoli casi d'uso, soprattutto in presenza di processi di apprendimento automatico che hanno un'incertezza statistica incorporata nelle loro definizioni (allegato III punto 5 lett. b). Infatti, la proposta di Regolamento sull'IA perno centrale il concetto di rischio per valutazione dell'affidabilità dei prestiti automatizzati, ma allo stesso tempo pone dei requisiti contradditori (...a eccezione dei sistemi di IA messi a servizio per uso da fornitori di piccole dimensioni).*

*In passato questi algoritmi hanno prodotti dei risultati socialmente discriminatori, come la carta di credito Apple sviluppata dalla banca d'investimento Goldman Saches. Secondo le segnalazioni fatti dagli utenti le donne, pur avendo la stessa ricchezza e credito rispetto agli uomini, ricevevano meno credito, pur se i programmatori non avessero avvertito questa problematica. La domanda è se il veto è posto ai fornitori di piccole dimensioni ma i chi controlla "la sicurezza" imposta delle big-tech?*

*La valutazione di un sistema di IA ad alto rischio è posta soprattutto al concetto di danno e impatto negativo verso le persone, infatti questo concetto viene rafforzato nella definizione ai sensi dall'art. 3 lettera c della proposta recante raccomandazioni alla Commissione su un regime di responsabilità civile per l'intelligenza artificiale (2020/2014(INL)) " "alto rischio": un potenziale significativo in un sistema di IA che opera in modo autonomo di causare danni o pregiudizi a una o più persone in modo casuale e che va oltre quanto ci si possa ragionevolmente aspettare; l'importanza del potenziale dipende dall'interazione tra la gravità dei possibili danni o pregiudizi, dal grado di autonomia decisionale, dalla probabilità che il rischio si concretizzi e dalla modalità e dal contesto di utilizzo del sistema di IA" , in cui cita l'interazione tra gravità dei possibili danni e dalla probabilità che il rischio si concretizzi. La Commissione europea, in questa proposta, ampia i criteri di valutazione dei sistemi di IA (art. 7, proposta di Regolamento sull'IA) dove il danno non ricade solo sulla salute, la sicurezza o l'impatto negativo sui diritti fondamentali ma ha anche un impatto negativo:*

- *sui diritti fondamentali;*
- *sulla salute;*
- *sulla sicurezza;*
- *sulla possibilità di sottrarsi a risultati discriminatori per motivi pratici o giuridici;*
- *sull'equità di potere, conoscenza, situazione economica o sociale o età.*

[5] Cfr. art. 7 comm.2 della proposta di Regolamento sull'IA "...se un sistema di IA presenti un rischio di danno per la salute e la sicurezza ..."

*Alla luce di ciò, la Proposta ha strutturata una classe di sistemi ad alto rischio come riportata alla tabella 1.*

| | |
|---|---|
| *Identificazione biometrica e classificazione delle persone fisiche* | *Utilizzato per l'identificazione biometrica a distanza "in tempo reale" e "a posteriori" – qualora non sia una pratica vietata;* |
| *Gestione e funzionamento delle infrastrutture critiche* | *Sistemi di IA utilizzati come componenti di sicurezza nel traffico stradale e nella fornitura di acqua, gas, riscaldamento ed elettricità;* |
| *Istruzione e formazione professionale* | *Sistemi di IA per (i) determinare l'accesso o l'assegnazione di soggetti a istituti di formazione; o (ii) valutare gli studenti negli istituti di istruzione (anche per i test di ammissione);* |
| *Occupazione, gestione dei lavoratori e accesso al lavoro autonomo* | *Sistemi di IA per (i) la ricerca e selezione del personale, ad esempio per posizioni lavorative aperte, screening/selezione dei candidati, valutazione di colloqui/test; o (ii) decisioni su promozione, cessazione del rapporto di lavoro, monitoraggio delle prestazioni;* |
| *Accesso a prestazioni e servizi pubblici e a servizi privati essenziali e fruizione degli stessi* | *Sistemi di IA utilizzati per: (i) determinare l'idoneità per benefici e servizi pubblici e le attività connesse; (ii) valutare l'affidabilità creditizia o stabilire punteggi di credito; o (iii) inviare, o stabilire la priorità per l'invio di servizi di emergenza;* |
| *Attività di contrasto* | *Sistemi di IA utilizzati per (i) valutare il rischio che una persona fisica commetta (o commetta nuovamente) un reato o il rischio per le potenziali vittime di reati; (ii) effettuare poligrafi o altro per rilevare lo stato emotivo di una persona fisica; (iii) rilevare deep fake; (iv) valutare l'affidabilità delle prove nelle indagini/procedimenti penali; (v) prevedere il (ri)verificarsi di reati così come l'individuazione, l'indagine o il perseguimento di reati basati sulla profilazione delle persone fisiche; o (vi) assistere con l'analisi del crimine e la ricerca di serie di dati per identificare modelli/relazioni;* |

| | |
|---|---|
| *Gestione della migrazione, dell'asilo e del controllo delle frontiere* | *Sistemi di IA usati da autorità pubbliche per (i) effettuare poligrafi o rilevare in altro modo lo stato emotivo di una persona; (ii) valutare i rischi di persone che intendono entrare in uno Stato membro per quanto riguarda, ad esempio, la sicurezza, l'immigrazione irregolare, la salute; (iii) verificare l'autenticità dei documenti di viaggio;*<br>*o (iv) assistere nell'esame delle domande di asilo, visti e permessi di soggiorno e l'ammissibilità;* |
| *Amministrazione della giustizia e processi democratici* | *Sistemi di IA per assistere le autorità giudiziarie nella ricerca/ interpretazione dei fatti e della legge, e nell'applicazione della legge adi una serie concreta di fatti.* |

*Tabella 1- Sistemi di IA ad alto rischio elencati nell'allegato III della proposta di Regolamento sull'IA.*

*Questo elenco potrà essere aggiornato dalla Commissione europea (ai sensi dell'art. 7 comma 1, proposta di Regolamento sull'IA) oppure integrato da altri sistemi laddove i casi di applicazione ed i rischi siano omogenei alla categoria di appartenenza.*

*La tabella qui di seguito riportata è la rappresentazione dei rischi legati IA prodotti che utilizzano parzialmente o completamente software IA e le soluzioni adottate in base alla loro classificazione.*

| *Rischi per i diritti fondamentali* | *Tipi di prodotti classificati in base al rischio* | *Soluzioni* |
|---|---|---|
| *totale (art. 5)* | *prodotti suscettibili di:*<br>*-causare o poter causare danni fisici o psicologici manipolando il comportamento umano per aggirare il libero arbitrio degli utenti;*<br>*- di imporre il c.d. 'social scoring' da parte o per conto delle autorità pubbliche che può portare a un trattamento dannoso o sfavorevole.* | *L'uso di questi prodotti è proibito.* |

| | | |
|---|---|---|
| *da totale ad alto in base all'uso che si fa del prodotto (art. 5)* | *i sistemi di identificazione biometrica a distanza "in tempo reale" in spazi accessibili al pubblico usate dalle forze dell'ordine rientrano in questa categoria.* | *Questo tipo di prodotto, sebbene proibito in linea generale, in casi eccezionali e sotto il controllo di un'autorità, è soggetto alle stesse regole stabilite per i prodotti ad alto rischio.* |
| *alto (art.6)* | *prodotti*<br>*- già soggetti alla legislazione europea di cui all'Allegato II del Regolamento ed alla valutazione di conformità da parte di terzi in vista dell'immissione sul mercato o della messa in servizio di tale prodotto IA sensi della stessa legislazione elencata nell'Allegato II;*<br>*- elencati nell'Allegato III del regolamento nei seguenti settori:*<br>*identificazione e categorizzazione biometrica delle persone fisiche; gestione e funzionamento delle infrastrutture critiche; istruzione e formazione professionale; occupazione, gestione dei lavoratori e accesso al lavoro autonomo; accesso e fruizione di servizi privati essenziali e di servizi e benefici pubblici; applicazione della legge; gestione della migrazione, dell'asilo e dei controlli* | *Al fine di mitigare l'alto rischio, il Regolamento:*<br>*- stabilisce che prima di immettere sul mercato tali prodotti, siano rispettati le seguenti condizioni:*<br>*• adozione di un sistema di gestione del rischio per valutare e contrastare il rischio;*<br>*• gestione dei dati affinché siano di alta qualità. Più alta è la qualità dei dati, più bassi sono i rischi per i diritti fondamentali e il rischio di esiti discriminatori;*<br>*• adozione di documentazione tecnica che fornisca tutte le informazioni necessarie per valutare la conformità del sistema IA requisiti e per consentire alle autorità di valutare tale conformità;*<br>*• conservazione della documentazione tramite registri (file di log) per garantire la tracciabilità dei risultati;*<br>*• trasparenza nei rapporti con gli utenti e doveri di informarli;*<br>*• supervisione umana per controllare e minimizzare il rischio;*<br>*• garantire robustezza, sicurezza e accuratezza nella gestione dell'IA affinché i risultati non siano corrotti o soggetti ad errori, difetti o incongruenze;* |

| | | |
|---|---|---|
| | *alle frontiere; amministrazione della giustizia e processi democratici.* | *- obbliga i fornitori, i distributori e gli utenti (artt. 16-29) a rispettare le condizioni di cui al punto precedente.*<br>*- stabilisce un sistema di standard, valutazione della conformità, certificati e registrazione per accertare e certificare l'adozione delle condizioni di cui al primo punto.* |
| *medio (art. 52)* | *-prodotti che interagiscono con le persone fisiche.* | *le persone fisiche devono essere informate che stanno interagendo con una IA.* |
| | *-prodotti per il riconoscimento delle emozioni o la categorizzazione biometrica.* | *le persone fisiche devono essere informate che sono esposte a un sistema di riconoscimento delle emozioni o di categorizzazione biometrica.* |
| | *-prodotti che generano o manipolano immagini, -contenuti audio o video che assomigliano sensibilmente a persone, oggetti, luoghi o altre entità o eventi esistenti e che potrebbero falsamente apparire a una persona come autentici o veritieri.* | *le persone fisiche devono essere informate che il contenuto è stato generato o manipolato artificialmente.* |

*Tabella 2-Elenco dei rischi e soluzioni adottate in base alla classificazione dei prodotti - fonte Altalex.com*

## Rischio medio

*A questa categoria di sistemi di IA appartengono sia quelli che presentano la tecnologia ad alto rischio e sia quelli che ne sono sprovvisti. L'elemento distintivo di questa categoria è* **la trasparenza** *nell'informare l'utente che stanno interagendo con un sistema di IA, a meno che non risulti evidente dalle circostanze e dal contesto di utilizzo.*

*Simile obbligo di trasparenza è previsto dalla norma anche per quanto riguarda i deepfake (sistemi di IA che generano o manipolano immagini o contenuti audio o video che assomigliano notevolmente a persone, oggetti, luoghi o altre entità o eventi esistenti e che potrebbero apparire falsamente autentici o veritieri per una persona) i*

*quali sono tenuti a rendere noto che il contenuto è stato generato o manipolato artificialmente.*

*A tali sistemi di IA è richiesto anche l'obbligo di informare l'utente a riguardo: dell'identità e dati di contatto del fornitore, le caratteristiche delle prestazioni, le misure di sorveglianza umana, le eventuali modifiche apportate al sistema di IA, la durata prevista e le misure di manutenzione.*

## *L'impiego dei sistemi di IA nei servizi pubblici e pubblica amministrazione*

*Lo scopo principale dell'UE è quello di diventare leader a livello mondiale nello sviluppo e nella diffusione di un'IA all'avanguardia, etica e affidabile, con l'intento di incentivare un approccio antropocentrico a livello globale.*

*Il piano concordato dall'UE deve essere una spinta verso questa nuova tecnologia per gli Stati membri a sviluppare le strategie nazionali di IA. Infatti, dovrebbe delineare gli obiettivi di investimento e le misure di attuazione e collaborare con gli Stati membri al fine di sviluppare degli indicatori comuni per il monitoraggio e l'analisi della strategia messa in atto.*

*Attraverso il servizio istituito dalla Commissione europea detto* **IA Watch4** *è possibile monitorare lo sviluppo, l'adozione e l'impatto dell'IA in Europa. Potenziato per lo sviluppo dell'IA nel settore pubblico, prevede l'analisi dell'uso e dell'impatto dell'IA relativamente ai servizi, raccogliendo informazioni sulle iniziative in corso degli Stati membri dell'UE e sviluppando una metodologia per identificare rischi e opportunità, driver e barriere all'uso dell'IA nei servizi pubblici.*

*Il settore pubblico in questa circostanza assume un doppio ruolo: quello di utilizzatore, come "utente", dell'IA e quello di "regolatore" nel definire le regole e gli obiettivi politici per un suo sviluppo etico come definito nel "Pacchetto digitale "dell'UE del 19 febbraio 2020.*

*Gli obiettivi principali del servizio IA Watch sono i seguenti:*

1. *Raccogliere informazioni sulle iniziative degli Stati membri dell'UE sull'uso dell'IA nei servizi pubblici;*
2. *sviluppare una proposta metodologica per identificare rischi e opportunità, fattori trainanti e ostacoli all'uso dell'IA nella fornitura di servizi pubblici e come valutarne gli impatti;*
3. *definire linee guida e una tabella di marcia generica per l'implementazione dell'IA nei servizi pubblici.*

*Per consentire il monitoraggio dello sviluppo dell'IA in Europa c'è bisogno che ogni Stato membro svolga tre attività principali:*

1. *ricerca esplorativa e paesaggistica dell'uso dell'IA a supporto dei servizi pubblici negli Stati membri dell'UE attraverso la mappatura e studi dei casi;*

2. *proposta di un quadro metodologico per la valutazione dell'impatto sociale ed economico, dopo l'identificazione dei servizi pubblici più promettenti che utilizzano l'IA;*
3. *progettazione di una roadmap di implementazione generica per l'uso dell'IA nei servizi pubblici, con linee guida e raccomandazioni basate sull'evidenza.*

## *Tipologia e finalità dell'IA nei servizi pubblici*

*Al fine di valutare quali tipologie di IA sono utilizzate in ambito europeo, è importante fare una classificazione, in modo da raggrupparne i diversi usi, come riportato di seguito:*

- ***Elaborazione audio**: queste applicazioni sono in grado di rilevare e riconoscere suoni, musica e altri input audio, incluso il parlato, consentendo così il riconoscimento delle voci e la trascrizione delle parole pronunciate.*
- ***Chatbot, assistenti digitali intelligenti, agenti virtuali e sistemi di raccomandazione**: questa tipologia di intelligenza artificiale include assistenti virtualizzati o "bot" online attualmente utilizzati negli ambienti CRM, sia nel settore privato che in quello pubblico, non solo per fornire consigli generici ma anche raccomandazioni relative al comportamento agli utenti.*
- ***Robotica cognitiva, automazione dei processi e veicoli connessi e automatizzati**: il tratto comune di queste tecnologie di intelligenza artificiale è l'automazione dei processi, che può essere ottenuta tramite hardware robotizzato (come arti prostatici o apparecchiature per la chirurgia di precisione) o software (o seguendo regole basate su apprendimento o approcci misti). È incluso anche l'uso di veicoli senza equipaggio per fornire servizi (ad esempio per la mobilità indipendente delle persone disabili).*
- ***Visione artificiale e riconoscimento dell'identità**: le applicazioni di intelligenza artificiale di questa categoria utilizzano una qualche forma di riconoscimento di immagini, video o facciali per ottenere informazioni sull'ambiente esterno e/o sull'identità di persone o oggetti specifici.*
- ***Sistemi esperti e basati su regole, processi decisionali algoritmici**: inizialmente lo sviluppo di questi sistemi di IA faceva parte di un'unica applicazione, adesso sono stati uniti a gestionali per facilitare o automatizzare i processi decisionali di rilevante importanza non solo nel settore privato ma anche in quello pubblico.*
- ***Gestione della potenziale conoscenza dall'intelligenza artificiale**: l'elemento comune qui è la capacità incorporata dell'intelligenza artificiale di creare una raccolta ricercabile di descrizioni di casi, testi e altri approfondimenti da condividere con gli esperti per ulteriori analisi attraverso delle ricerche mirate.*
- ***Machine Learning, Deep Learning**: mentre quasi tutte le altre categorie di intelligenza artificiale utilizzano una qualche forma di apprendimento*

*automatico, questa categoria si riferisce a soluzioni di intelligenza artificiale che non sono adatte alle altre classificazioni.*

- ***Elaborazione del linguaggio naturale, estrazione di testo e analisi del parlato**: queste applicazioni sono in grado di riconoscere e analizzare il parlato, il testo scritto e comunicare.*
- ***Analisi predittiva, simulazione e visualizzazione dei dati**: queste soluzioni apprendono da grandi set di dati per identificare modelli nei dati che vengono di conseguenza utilizzati per visualizzare, simulare o prevedere nuove configurazioni.*

| TIPLOGIA DI IA/ | SG | P | DIF | OP | EA | P | AC | SA | DIV | ED | Tota |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Processi Audio | 6 | | | 1 | | | | 1 | | | 8 |
| Chatbots, Assistente Digitale Intelligente, Agente Virtuale e Sistemi di | 36 | 1 | | | 7 | | | 7 | 1 | | 52 |
| Robotica cognitive, Automazione dei processi e veicoli | 1 | 1 | | 3 | 6 | | 1 | 4 | | | 16 |
| Visione artificiale, Riconoscimento | 5 | | 1 | 3 | 9 | 3 | 3 | 4 | 1 | | 29 |
| Sistemi esperti basati su regole, Processo decisionale algoritmico | 5 | 3 | | 5 | 4 | | 2 | 8 | | 2 | 29 |
| Gestione della conoscenza basata | 4 | 2 | | 2 | | | | 1 | 2 | 1 | 12 |
| Machine Learning, Deep Learning | 4 | 1 | | 4 | 3 | | 1 | 2 | | 2 | 17 |
| Elaborazione del linguaggio naturale, Estrazione di testo e | 10 | 1 | 1 | 2 | 1 | | 1 | 2 | | 1 | 19 |
| Analisi predittiva, Simulazione e visualizzazione dei | 4 | 4 | | 1 | 9 | | 6 | 12 | | 1 | 37 |
| Analisi della sicurezza, Intelligence sulle | 1 | 1 | 2 | 6 | 1 | | | | | | 11 |
| Totale | 76 | 14 | 4 | 27 | 40 | 3 | 14 | 41 | 4 | 7 | 230 |

*Tabella 3- Analisi delle tipologie di IA utilizzate nei vari settori. Leggenda dei settori SGP servizi pubblici generali, PC Protezione civile, DIF Difesa, EA Economia e Affari, OPS Ordine pubblica e sicurezza, PA Protezione dell'ambiente, AC Abitazione e comunità, SAL Salute, DIV Divertimenti, EDU Educazione*

*Da una prima analisi della tabella 3 risulta che solo due classi sono maggiormente utilizzate dalla pubblica amministrazione nello specifico l'interazione live con i "clienti" per la fornitura di supporto online tramite chatbot e simili – e lo sfruttamento dei dati disponibili mediante strumenti di visualizzazione, simulazione e predizioni; tuttavia anche se sono servizi disgiunti dal supporto immediato al processo decisionale essi possono viaggiare in parallelo, come ad esempio quelli rivolti a disegnare scenari che migliorino la comprensione umana in questioni complesse, sociali o organizzative.*

*Si può osservare che le combinazioni più sorprendenti sono:*

- *Assistenti virtuali nei servizi pubblici generali (con 36 casi)*
- *Analisi predittiva (per la creazione di scenari) negli ambienti sanitari, con 12 casi.*

*Le altre combinazioni che sono degne di nota riguardano l'uso di Computer Vision e Predictive Analytics nel dominio degli affari economici, entrambi con nove casi.*

*L'impiego dell'IA in questi mesi di crisi COVID -19 ha trasformato gli studi teorici in soluzioni immediatamente applicate sul campo, ne è esempio il caso del software InferRead CT Lung COVID-19 finanziato dall'UE nell'ambito di un progetto pilota in cui i medici non sono più soli in questa lotta perché possono ricevere informazioni attraverso interoperabilità di banche dati di altri 11 ospedali e possono anche fornire servizi sanitari digitali di alta qualità.*

*In alcuni casi come quello sopra descritto si può dire che iniziative ad alto impatto sociale hanno evidenziato una visione preventiva da parte dei governi che le hanno finanziate, mentre alcuni sono rimasti dei progetti sviluppati in soluzioni legati alla pura ricerca ma senza alcun potenziale utilizzo nei servizi pubblici.*

*Il vento dell'innovazione tecnologica, seppur con tempistiche non sempre celeri e le immense difficoltà dovute da adattamenti organizzativi, ha investito inesorabilmente la PA. La sua presenza è stata da tempo sponsorizzata dal* **Codice dell'Amministrazione Digitale (CAD) D.L. del 7 marzo 2005 n 82 che esprime, dal lontano 2005**, *all'art. 2 "la disponibilità, la gestione, l'accesso. La trasmissione, la conservazione e la fruibilità dell'informazione in modalità digitale… utilizzando con le modalità più appropriate, le tecnologie dell'informazione e della comunicazione…"*

*In particolare, il successivo art. 12 precisa, al comma 1, che "[l]e pubbliche amministrazioni nell'organizzare autonomamente la propria attività utilizzano le tecnologie dell'informazione e della comunicazione per la realizzazione degli obiettivi di efficienza, efficacia, economicità, imparzialità, trasparenza, semplificazione e partecipazione nel rispetto dei principi di uguaglianza e di non discriminazione, nonché per l'effettivo riconoscimento dei diritti dei cittadini e delle imprese […]".*

*Ad ulteriore conferma, l'art. 3-bis della legge n. 241/90 sul procedimento amministrativo promuove l'uso della telematica, sia tra i rapporti interni alle pubbliche amministrazioni che nei rapporti con i privati, al fine di conseguire "maggiore efficienza".*

*Ad ogni modo, l'introduzione dei sistemi di IA apporta* **modifiche incrementali all'erogazione del servizio pubblico***, mentre altri mirano ad avere cambiamenti molto più dirompenti,* **ridisegnando le pratiche di lavoro esistenti***. Allo stesso tempo, alcuni dei casi di più radicale innovazione basata sull'intelligenza artificiale sollevano* **preoccupazioni e timori da parte di cittadini e autorità di regolamentazione***, poiché possono ridefinire le relazioni di potere all'interno dell'area della governance e portare nuovi squilibri di rischio nei contesti democratici delle società europee.*

## *Casi di utilizzo dei sistemi di IA nel settore pubblico*

*A livello Europeo è stato scelto di analizzare, per le motivazioni suddette, alcuni casi controversi di utilizzo dell'IA, che sono stati sospesi o sono al vaglio della magistratura, per motivi etici, legali e sociali. Alcuni di essi sono rappresentati e descritti nelle pagine successive. Questi casi evidenziano che l'introduzione e l'uso di sistemi di intelligenza artificiale nelle organizzazioni e negli ambienti del settore pubblico, non essendo approcciati solo per i requisiti tecnologici ma, coinvolgendo la percezione dei cittadini e dei dipendenti pubblici che ne fanno uso, sono da considerarsi un elemento cruciale per un uso sostenibile nei servizi e nelle politiche pubbliche.*

## *Sistema preventivo, Belgio*

*Nel 2014, l'Agenzia fiamminga per l'infanzia e la famiglia (Kind en Gezin) ha sviluppato un sistema di intelligenza artificiale che consente previsioni più accurate utili a rilevare i servizi di assistenza diurna che richiedono ulteriori ispezioni.*

*Queste ispezioni consentono di mantenere l'alta qualità dei servizi di asilo nido e di migliorare il benessere dei bambini. L'Agenzia per l'infanzia e la famiglia non effettua direttamente i controlli, ma collabora con l'Unità regionale di ispezione sanitaria del Dipartimento del Welfare, della Sanità pubblica e della famiglia.*

*Tuttavia, la capacità di condurre tutte le ispezioni è limitata. Per un lungo periodo di tempo, c'è stato un interesse nel capire come ottimizzarla. L'uso dei dati era stato considerato un modo per* **migliorare le pratiche di ispezione esistenti e ottimizzare la scarsa quantità di ispettori***. Il sistema predittivo sviluppato utilizza un metodo di apprendimento automatico supervisionato (regressione logistica e XGBoost) per analizzare i vari dati interni ed esterni dell'Unità sanitaria. Le raccomandazioni combinate del sistema predittivo associate all'esperienza e le competenze del personale sanitario consentono interventi più mirati e basati sulle situazioni esistenti.*

*Ora, il sistema è apprezzato dai dipendenti pubblici, anche se ancora bisogna lavorare per convincerli del valore aggiunto che può dare questo modello. In particolare, il personale doveva essere convinto che* ***l'uso del modello fosse inteso a responsabilizzarli, non a sostituire le loro competenze o a controllare il loro lavoro****. Alla fine, la combinazione delle dimostrazioni statistiche sulla validità del sistema e dell'enfasi sul sostegno ai lavoratori umani, ha ulteriormente migliorato l'accettazione e il sostegno degli utenti finali dei servizi del settore pubblico.*

*Un'altra importante intuizione di questo caso è la necessità per le organizzazioni pubbliche di fornire una manutenzione continua del modello e dei dati sottostanti al fine di rendere permanente l'adozione dell'IA. Il sistema di intelligenza artificiale ha avuto una costante manutenzione e miglioramento del modello al fine di garantirne l'accuratezza e l'affidabilità: se questa manutenzione fosse ignorata, l'accuratezza del modello potrebbe diminuire, il che ridurrebbe la fiducia nel modello e in altri futuri progetti.*

## *Servizi pubblici automatizzati a Trelleborg, Svezia*

*Nel comune di Trelleborg fin dal 2016, le nuove tecnologie vengono utilizzate per automatizzare varie decisioni di assistenza sociale. Questo è stato il primo comune a utilizzare Robotic Process Automation (RPA) per gestire le varie applicazioni di assistenza sociale. Al momento, il sistema decisionale automatizzato è in grado di elaborare le domande di assistenza domiciliare, indennità di malattia, indennità di disoccupazione e tasse ed è stato considerato un esempio di successo da seguire.*

*In precedenza, i dipendenti dovevano valutare manualmente le domande ricevute, il che richiedeva tempi e costi considerevoli. Con più di 300 richieste di prestazioni sociali nel comune ogni mese, i cittadini a volte hanno dovuto aspettare in media 8 giorni per avere una risposta sulle loro prestazioni sociali, a volte anche fino a 20 giorni. A causa di questo tempo di attesa, i cittadini avrebbero spesso contattato il dipartimento per sapere l'esito, aumentando ulteriormente il carico di lavoro del personale.* ***La decisione di utilizzare l'IA per migliorare il processo è stata presa per limitare i tempi di attesa e anche varie preoccupazioni legate ai ritardi nei pagamenti ai cittadini****. L'automazione dei servizi di welfare non è stata possibile fino a quando non è stato reso disponibile un processo online per la presentazione delle domande di welfare. Nel 2015, Trelleborg è stato il primo comune svedese a digitalizzare l'amministrazione per le prestazioni sociali. Ad oggi il 75% dei cittadini utilizza la piattaforma online per accedere ai pagamenti assistenziali, che ha consentito l'acquisizione di dati e informazioni significativi per automatizzare questo processo con la nuova tecnologia. Senza i dati provenienti dal portale self-service non sarebbe stato possibile automatizzare questi processi.*

*Grazie al processo di automazione,* ***i tempi di attesa per i cittadini sulle loro domande di welfare sono stati notevolmente ridotti****. È stato detto che in molti casi i tempi di gestione per le persone in situazioni economicamente vulnerabili sono stati ridotti da 10 a 1 giorno. Inoltre, due dipendenti dell'amministrazione Trelleborg sono stati riassegnati per dedicare più tempo ad altre attività di valore aggiunto come la gestione di casi più complessi. Da un primo studio sull'uso dell'automazione ha rilevato un atteggiamento positivo da parte del personale nei confronti dell'uso di queste tecnologie ha permesso un lavoro più efficace e ha portato una certezza giuridica.*

*Nonostante questi effetti positivi, ci sono state delle preoccupazioni per l'uso dell'automazione. Durante la fase di implementazione, molti assistenti sociali erano restii a utilizzare il sistema per* ***paura di perdere il lavoro o di trasferire compiti sociali sensibili ai computer****. In altri comuni svedesi dove è stato adottato il sistema di IA Trelleborg hanno incontrato della resistenza da parte di alcuni membri del loro personale locale, alcuni dei quali sono stati persino portati a dimettersi. Occorre fare attenzione che nel momento in cui tutti i processi sono automatizzati online si rischia di escludere la classe di cittadino più vulnerabile e ciò renderebbe più difficile la valutazione delle loro esigenze*

*Mentre il sistema di intelligenza artificiale ha consentito di automatizzare varie decisioni relative IA benefici sociali, molti altri processi del comune di Trelleborg funzionano ancora come in un sistema burocratico tradizionale. Ci sono ancora molti processi cartacei all'interno dell'organizzazione che potrebbero portare a una doppia documentazione e processi inefficienti, così come esistenti software con interfacce e livelli di usabilità molto scarsi.*

## *Profilazione dei disoccupati, Polonia*

*Già nel 2012 il Ministero polacco del Lavoro e delle Politiche Sociali (MLSP) ha iniziato a lavorare alla riforma dei 340 uffici del lavoro (PUP - Powiatowe Urzędy Pracy), incaricati di analizzare le tendenze e sostenere lo sviluppo del mercato del lavoro. L'urgenza della riforma è stata sottolineata dalla percezione generale che i PUP siano inefficienti, a corto di personale e inadatti ad affrontare le sfide poste dal moderno mercato del lavoro.*

*Con tale riforma il MLSP ha individuato possibili soluzioni che garantiscano una più efficiente allocazione del budget. Con questa prospettiva vi è ricorso a un sistema di profilazione automatizzato per la disoccupazione e ha rilevato un metodo di erogazione dei servizi moderno, efficiente in termini di costi ed individualizzato.*

*Il processo di profilazione automatizzata divide i disoccupati in tre categorie, tenendo conto di una serie di caratteristiche individuali. L'assegnazione a una specifica categoria determina i tipi di programmi per i quali un beneficiario è idoneo (ad esempio inserimento lavorativo, formazione professionale, apprendistato, indennità di*

*attivazione). Il sistema si basa sui dati raccolti durante un colloquio iniziale (es. età, sesso, disabilità e durata della disoccupazione) e un successivo test computerizzato che valuta 24 diverse dimensioni. L'assegnazione a uno dei tre gruppi di profili indica il livello necessario di supporto e l'onere delle risorse. È importante sottolineare che in ogni caso la categorizzazione produce decisioni binarie tanto da far cambiare la vita: sostegno statale o mancanza di esso. Però,* **l'uso di questo sistema di IA ha ricevuto delle critiche sia all'interno dell'organizzazione che all'esterno**. *In primo luogo,* **l'algoritmo dell'applicazione è opaco** *poiché ai cittadini non viene detto del punteggio ricevuto né di come questo punteggio sia stato determinato. Inoltre, l'idea alla base del meccanismo di profilazione era di servire esclusivamente come strumento di consulenza, mantenendo un essere umano a esprimere il proprio giudizio sull'output della categoria di lavoro.*

*Sorprendentemente il risultato prodotto dallo studio, meno di 1 su 100 decisioni prese dall'algoritmo, è stato messo in discussione dai responsabili d'ufficio. Anche se l'applicazione restituisce un risultato abbastanza preciso, vi sono altri motivi che mettono in discussione questo sistema: la mancanza del tempo per analizzare i risultati e una presunta obbiettività dei processi; ricavando una fragilità di rapporto tra processi e dipendente a causa dello scemare del meccanismo consultivo.*

*Molti disoccupati si sono rivolti ai tribunali amministrativi, sostenendo che la categorizzazione era ingiusta.* **L'Ufficio supremo dei controlli ha effettuato un controllo approfondito dei PUP e ha giudicato l'inefficacia del sistema di profilazione e malevole il suo potere di portare ad azioni discriminatorie.** *Infine, il Commissario per i diritti umani ha presentato una denuncia formale alla Corte costituzionale per una questione procedurale, e quest'ultima ha stabilito che lo strumento di profilazione era incostituzionale. A partire dal 14 giugno 2019, il sistema è stato ufficialmente smantellato dal governo.*

## *VeriPol, Spagna*

*La presentazione di falsi rapporti da parte della polizia è abbastanza comune in Spagna, soprattutto per reati di basso livello. Questa pratica è considerata piuttosto problematica, in quanto può comportare conseguenze significative per gli individui.*

*Recentemente la polizia nazionale spagnola ha adottato il sistema VeriPol IA per rilevare i falsi rapporti della polizia. Il sistema è stato progettato per essere integrato nel sistema informativo esistente della polizia nazionale spagnola chiamato SIDENPOL, consentendo un uso più semplice e integrato nelle pratiche di lavoro esistenti.* **Il suo sviluppo è stato il risultato di un progetto di collaborazione tra l'Università di Cardiff, l'Università Carlo III di Madrid e la polizia nazionale spagnola**. *La banca dati dei verbali della polizia è stata messa a disposizione ai ricercatori e sono state utilizzate 1122 segnalazioni, di cui 534 vere e 588 false.*

*VeriPol sfrutta una* **combinazione di elaborazione del linguaggio naturale e algoritmi di classificazione dell'apprendimento automatico**, *in grado di stimare la probabilità di falsi rapporti di polizia con una precisione assoluta. Oltre a ciò, il sistema consente anche di approfondire le* **differenze tra rapporti di polizia falsi e veri**. *Ad esempio, lo studio pilota ha scoperto che i rapporti falsi della polizia hanno maggiori probabilità di includere dichiarazioni più brevi, incentrate sugli oggetti rubati e prive di dettagli.*

*In seguito allo sviluppo, il sistema è stato testato sia dal dipartimento di polizia di Malaga che quello di Murcia. Questi progetti sono stati considerati un grande successo visto l'aumento del numero di false segnalazioni; lo studio sta proseguendo per ampliare altre funzionalità in grado di rilevare altre forme di criminalità.*

**Ora il sistema è stato implementato per essere utilizzato da tutti i dipartimenti della polizia nazionale spagnola**. *L'impatto previsto dell'uso del sistema è sia quello di rilevare tempestivamente le false segnalazioni, lasciando più risorse di polizia disponibili per concentrarsi su altri compiti e segnalazioni, e allo stesso tempo è un ottimo messo per dissuadere le persone dal presentare false dichiarazioni. Un ulteriore vantaggio del sistema è quello di acquisire maggiori informazioni su come le persone mentono agli agenti di polizia e di acquisire maggiori conoscenze nell'individuazione di rapporti di polizia veri e falsi.*

## *Italia, utilizzo dei sistemi di IA nella Pubblica Amministrazione*

*Tra i processi di IA troviamo quelle relative agli assistenti virtuali, quelle relative al supporto IA processi e quelle relative al supporto IA dati. Iniziamo dalle soluzioni impiegate come assistenti virtuali, un primo esempio è dato dal Comune di Solarino, che ha implementato un Ufficio Relazioni con il Pubblico detto "intelligente" detto AXEL; la sua funzione è di rispondere contemporaneamente a moltissime richieste, attraverso canali diversi che vanno dalla presenza in loco del cittadino alla chiamata o alla e-mail.*

*AXEL è un modello in grado di comprendere le domande poste dagli utenti nel linguaggio comune e rispondere o tramite e-mail o tramite SMS oppure direttamente dalla chat del sito istituzionale per arrivare a rispondere anche sui social network, su Skype e al telefono.*

*Un altro esempio di IA negli enti locali è "020202", assistente virtuale del comune di Milano, disponibile su WhatsApp 24 ore su 24, la sua funzione è quella di fornire ai cittadini informazioni sulla città. "020202" rientra tra le chatbots e permette di ottenere risposte immediate sull'emergenza Covid-19, sulla città più in generale, sulla sanità, su documenti prodotti dal Governo.*

*Per i sistemi di IA a supporto dei processi troviamo il sistema di tele monitoraggio per i malati affetti da broncopneumopatia cronica ostruttiva, facendo ricorso al machine*

*learning e vedendo coinvolti l'Unita di Sistemi di Elaborazione e Bioinformatica del Campus Bio-Medico di Roma e il Policlinico universitario Campus Biomedico di Roma.*

*Infine, l'impiego dei sistemi di IA a supporto dei dati vi sono molti casi della Pubblica Amministrazione. Tra queste, l'Azienda Gardesana Servizi Spa, l'Università di Verona, l'Istituto Superiore di Sanità, che stanno facendo ricorso al progetto INTCATCH, un progetto di innovazione e monitoraggio dei bacini idrici, che vede il Lago di Garda come il più importante sito di innovazione d'Europa; finanziato nell'ambito di Horizon2020, il più prestigioso programma di ricerca europeo, L'obiettivo è quello di mettere a punto, validare e replicare metodi e tecniche robotiche e biotecnologiche di monitoraggio e di gestione dei bacini idrografici.*

*Un altro esempio di IA nel supporto ai dati è* **il progetto DANTE H2020**[10]*, che vede coinvolti Ministero della Difesa e Comando generale Arma dei Carabinieri.*

*Lo scopo del progetto è di fornire soluzioni di analisi sempre più efficaci, efficienti, automatizzate e di sviluppare un sistema integrato per* **rilevare, recuperare, raccogliere e analizzare un gran numero di multimedia eterogenei e complessi e contenuto multilingue relativo al terrorismo***, anche tramite il deep web e le dark nets. È risaputo che lo scambio di denaro è alla base di tutte le attività terroristiche e quindi l'interruzione dell'attività di finanziamento del terrorismo è la base della lotta al terrorismo. Lo scopo finale è quello di monitorare e controllare tutto ciò che può rientrare nelle attività terroristiche, come le raccolte fondi online e le attività di propaganda.*

*Sicuramente* **l'uso più incisivo dell'IA è quello di associarlo insieme alla digitalizzazione "comune"** *per permettere di attuare e rispettare il principio di trasparenza nell'ambito delle procedure di acquisto, anche ai fini di* **prevenzione della corruzione***, soprattutto nell'ambito degli appalti pubblici se spetta al software di IA selezionare la migliore offerta pervenuta da più società partecipanti.*

*Ma se la P.A. avesse adottato un atto amministrativo avvalendosi di un sistema di IA, si sarebbero dovute conoscere le motivazioni per le quali siano state assunte tali decisioni. A tal riguardo, il Consiglio di Stato*[11] *ha affermato che "il meccanismo attraverso il quale si concretizza la decisione robotizzata (ovvero l'algoritmo) deve essere "conoscibile", secondo una declinazione rafforzata del principio di trasparenza, che implica anche quello della piena conoscibilità di una regola espressa in un linguaggio differente da quello giuridico. Tale conoscibilità dell'algoritmo deve essere garantita in tutti gli aspetti: dai suoi autori al procedimento usato per la sua elaborazione, al meccanismo di decisione, comprensivo delle priorità assegnate nella procedura valutativa e decisionale e dei dati selezionati come rilevanti. Ciò al fine di poter verificare che gli esiti del procedimento robotizzato siano conformi alle prescrizioni e alle finalità stabilite dalla legge o dalla stessa amministrazione a monte di tale procedimento e affinché siano chiare – e conseguentemente sindacabili – le*

*modalità e le regole in base alle quali esso è stato impostato… In secondo luogo, la regola algoritmica deve essere non solo conoscibile in sé, ma anche soggetta alla piena cognizione, e al pieno sindacato, del giudice amministrativo."*

# *La trasparenza comunicativa: il principio chiave che potrebbe scardinare le oscurità dei sistemi di Intelligenza Artificiale nell'applicazione del processo algoritmico automatico.*

*di Anna Capoluongo, Andrea Piccoli, Valentina Sapuppo*

## *Il problema definitorio del concetto di Intelligenza Artificiale*

*Nell'evoluzione tecnologica e tra i fattori trainanti della trasformazione digitale anche della Pubblica Amministrazione*[6] *si pone la sempre più pervasiva applicazione del*

[6] In Italia, il Codice Amministrazione Digitale (CAD) recepisce all'art. 50-ter***** il progetto relativo alla Piattaforma Digitale Nazionale Dati (PDND). La Piattaforma Digitale Nazionale Dati (PDND), già precedentemente introdotta nel Piano Triennale per l'Informatica 2017-2019, è stata affidata alla società PagoPA lo sviluppo e la diffusione della PDND. Su tale tema si è espresso anche il GPDP con Parere sullo schema di "Linee Guida sull'infrastruttura tecnologica della Piattaforma Digitale Nazionale Dati per l'interoperabilità dei sistemi informativi e delle basi di dati" del 16 dicembre 2021.

*****Cfr. Art. 50-ter (Piattaforma Digitale Nazionale Dati) 1. La Presidenza del Consiglio dei ministri promuove la progettazione, lo sviluppo e la realizzazione di una Piattaforma Digitale Nazionale Dati (PDND) finalizzata a favorire la conoscenza e l'utilizzo del patrimonio informativo detenuto, per finalità istituzionali, dai soggetti di cui all'articolo 2, comma 2, nonché la condivisione dei dati tra i soggetti che hanno diritto ad accedervi ai fini dell'attuazione dell'articolo 50 e della semplificazione degli adempimenti amministrativi dei cittadini e delle imprese, in conformità alla disciplina vigente. 2. La Piattaforma Digitale Nazionale Dati è gestita dalla Presidenza del Consiglio dei ministri ed è costituita da un'infrastruttura tecnologica che rende possibile l'interoperabilità dei sistemi informativi e delle basi di dati delle pubbliche amministrazioni e dei gestori di servizi pubblici per le finalità di cui al comma 1, mediante l'accreditamento, l'identificazione e la gestione dei livelli di autorizzazione dei soggetti abilitati ad operare sulla stessa, nonché' la raccolta e conservazione delle informazioni relative agli accessi e alle transazioni effettuate suo tramite. La condivisione di dati e informazioni avviene attraverso la messa a disposizione e l'utilizzo, da parte dei soggetti accreditati, di interfacce di programmazione delle applicazioni (API). Le interfacce, sviluppate dai soggetti abilitati con il supporto della Presidenza del Consiglio dei ministri e in conformità alle Linee guida AgID in materia interoperabilità, sono raccolte nel "catalogo API" reso disponibile dalla Piattaforma ai soggetti accreditati. I soggetti di cui all'articolo 2, comma 2, sono tenuti ad accreditarsi alla piattaforma, a sviluppare le interfacce e a rendere disponibili le proprie basi dati senza nuovi o maggiori oneri per la finanza pubblica. In fase di prima applicazione la Piattaforma assicura prioritariamente l'interoperabilità con (le basi di dati) di interesse nazionale di cui all'articolo 60, comma 3-bis e con e banche dati dell'Agenzie delle entrate individuate dal Direttore della stessa Agenzia. L'AgID, sentito il Garante per la protezione dei dati personali e acquisito il parere della Conferenza unificata, di cui all'articolo 8 del decreto legislativo 28 agosto 1997, n. 281, adotta linee guida con cui definisce gli standard tecnologici e criteri di sicurezza, di accessibilità, di disponibilità e di interoperabilità per la gestione della piattaforma nonché' il processo di accreditamento e di fruizione del catalogo API con i limiti e le condizioni di accesso volti ad assicurare il corretto trattamento dei dati personali ai sensi della normativa vigente. 2-bis. Il Presidente del Consiglio dei ministri o il Ministro delegato per l'innovazione tecnologica e la transizione digitale, ultimati i test e le prove tecniche di corretto funzionamento della piattaforma, fissa il termine entro il quale i soggetti di cui all'articolo 2, comma 2, sono tenuti ad accreditarsi alla stessa, a sviluppare le interfacce di cui al comma 2 e a rendere disponibili le proprie basi dati. 3. Nella Piattaforma Nazionale Digitale Dati non confluiscono i dati attinenti a ordine e sicurezza pubblici, difesa e sicurezza nazionale, polizia giudiziaria e polizia economico-finanziaria. 4. Con decreto adottato dal Presidente del Consiglio dei ministri entro sessanta giorni dalla data di entrata in vigore della presente disposizione, di concerto con il Ministero dell'economia e delle finanze e il Ministero dell'interno, sentito il Garante per la protezione dei dati personali e acquisito il parere della Conferenza Unificata di cui all'articolo 8 del decreto legislativo 28 agosto 1997, n. 281, è stabilita la strategia nazionale dati. Con la strategia nazionale dati sono identificate le tipologie, i limiti, le finalità e le modalità di messa a disposizione, su richiesta della Presidenza del Consiglio dei ministri, dei dati aggregati e anonimizzati di cui sono titolari i soggetti di cui all'articolo 2, comma 2, dando priorità ai dati riguardanti gli studenti del sistema di istruzione e di istruzione e formazione professionale ai fini della realizzazione del diritto-dovere all'istruzione e alla formazione e del contrasto alla dispersione scolastica e formativa. 5. L'inadempimento dell'obbligo di rendere disponibili e accessibili le proprie basi dati ovvero i dati aggregati e anonimizzati costituisce mancato raggiungimento di uno specifico risultato e di un rilevante obiettivo da parte dei dirigenti responsabili delle strutture competenti e comporta la riduzione, non inferiore al 30 per cento, della retribuzione di risultato e del trattamento accessorio collegato alla performance individuale dei dirigenti competenti, oltre al divieto di attribuire premi o incentivi nell'ambito delle medesime strutture. 6. L'accesso ai dati attraverso la Piattaforma Digitale Nazionale Dati non modifica la disciplina relativa alla titolarità del trattamento, ferme restando le specifiche responsabilità ai sensi dell'articolo 28 del Regolamento (UE) 2016/679 del Parlamento Europeo e del Consiglio del 27 aprile 2016 in capo al soggetto gestore della Piattaforma nonché' le responsabilità dei soggetti accreditati che trattano i dati in qualità di titolari autonomi del trattamento. 7. Resta fermo che i soggetti di cui all'articolo 2, comma 2, possono continuare a utilizzare anche i sistemi di interoperabilità già previsti dalla legislazione vigente. 8. Le attività previste dal presente articolo si svolgono con le risorse umane, finanziarie e strumentali disponibili a legislazione vigente."

*processo algoritmico automatico, nonché delle soluzioni basate sulle nuove tecnologie di Intelligenza Artificiale.*

*I contesti funzionali in cui sta maggiormente diffondendosi tale prassi possono, attualmente, ben inquadrarsi nell'adozione di tecnologie volte alla profilazione massiva degli utenti, all'elaborazione di processi predittivi in ambito finanziario, giuridico*[7] *et similia, alla manutenzione preventiva, al riconoscimento e all'indicizzazione e classificazione delle informazioni.*

*Tra gli altri, rileviamo altresì l'applicazione di sistemi di analisi algoritmica operanti sulle risultanze di estrazione della conoscenza dai Big Data, che consentono di analizzare la realtà attraverso altri occhi (per esempio, è nota l'adozione di sistemi di Intelligenza Artificiale orientati alla diagnosi precoce di malattie e di trattamento terapico in ambito sanitario - mediante l'applicazione di algoritmi di knowledge extraction, ovvero quella riguardante l'automation nei sistemi di trasporto, contestualizzati in una prospettica visione di IoT entro il progetto delle smart cities - mediante l'analisi di impatto dei dispositivi innovativi e la loro sostenibilità, anche economica).*

*A fare da sponda agli sviluppi stupefacenti derivanti dalle nuove applicazioni dei sistemi di Intelligenza Artificiale, che arrivano persino ad ipotizzare l'applicabilità alle macchine della teoria della mente e quindi dell'empatia*[8]*, si stagliano zone d'ombra imponenti, costituite, in primis, dalla opacità degli algoritmi.*

*Nonostante le inferenze causate dai bias, in grado di causare discriminazioni sociali importanti nonché realizzare potenziali violazioni dei diritti umani fondamentali*[9]*, tali storture appaiono di fondamentale importanza al fine di raddrizzare il tiro, tenuto conto anche dell'alto livello di applicazione della normativa volta alla protezione dei dati personali delle persone fisiche.*

*In assenza di una definizione unitaria, globalmente accettata, che descriva ciò che debba intendersi per Intelligenza Artificiale, rileviamo che la dottrina scientifica ha registrato il grande contributo apportato da S. Russell e P. Norvig, i quali, nell'opera Artificial intelligence: a modern approach fourth edition*[10]*, affermano che "la definizione può cambiare a seconda del tipo di approccio che si intende attribuire fondamentalmente secondo quattro modelli: a) pensando in modo umano; b) pensando razionalmente; c) agendo in modo umano; agendo razionalmente".*

*Nella Comunicazione sull'A.I. per l'Europa*[11]*, invece, la Commissione Europea ha proposto una definizione perfezionata, secondo la quale: "I sistemi di Intelligenza*

---

[7] Per un maggior approfondimento sui temi della giustizia predittiva, si rimanda a Sapuppo, V. *"Predictive. L'algoritmo che condanna", Salvis Juribus.*

[8] Chen, B., Vondrick, C. & Lipson, H., *Visual behavior modelling for robotic theory of mind*, Sci Rep 11, 424 (2021).

[9] Si veda Ponticiello R., *Data Protection 3.0. Cosa è successo e quali sono le prospettive nel post Covid*, NT+ Il Sole24Ore.

[10] Russel, S., Norvig, P., *Artificial intelligence: a modern approach,* Pearson, 2021.

[11] Comunicazione della Commissione al Parlamento Europeo, al Consiglio, al Comitato Economico e Sociale Europeo e al Comitato delle Regioni, *L'intelligenza artificiale per l'Europa*, Bruxelles, 25.4.2018 COM(2018) 237 final.

*Artificiale (IA) sono sistemi software (e possibilmente hardware) progettati da esseri umani e che, avendo ricevuto un obiettivo complesso, agiscono nel mondo reale o digitale percependo il loro ambiente attraverso l'acquisizione di dati, interpretando i dati strutturati o non strutturati raccolti, applicando il ragionamento alla conoscenza o elaborando le informazioni derivate da questi dati e decidendo le migliori azioni da intraprendere per raggiungere l'obiettivo dato. I sistemi di Intelligenza Artificiale possono utilizzare regole simboliche o apprendere un modello numerico. Possono anche adattare il loro comportamento analizzando come l'ambiente è influenzato dalle loro azioni precedenti".*

*Affiancando tali definizioni a quella avanzata dalla Proposta di Regolamento sull'Intelligenza Artificiale[12], che all'articolo 3, par. 1 definisce "sistema di Intelligenza Artificiale" (sistema di IA), un software sviluppato con una o più delle tecniche e degli approcci elencati nell'allegato I, che può, per una determinata serie di obiettivi definiti dall'uomo, generare output quali contenuti, previsioni, raccomandazioni o decisioni che influenzano gli ambienti con cui interagiscono", potrebbe apparire più evidente e chiara la definizione del perimetro entro cui tali sistemi si posizionano.*

*In Italia, anche il Consiglio di Stato si è cimentato in questo esercizio definitorio. Con la sentenza n. 7891 del 25.11.2021[13], infatti, il Consiglio di Stato ha dimostrato un lodevole sforzo interpretativo, distinguendo tra le nozioni di intelligenza artificiale e di algoritmo. Secondo i Giudici di Palazzo Spada, infatti, "non v'è dubbio che la nozione comune e generale di algoritmo riporti alla mente* **semplicemente una sequenza finita di istruzioni, ben definite e non ambigue, così da poter essere eseguite meccanicamente e tali da produrre un determinato risultato.** *Nondimeno si osserva che la nozione, quando è applicata a sistemi tecnologici, è ineludibilmente collegata al concetto di automazione ossia a sistemi di azione e controllo idonei a ridurre l'intervento umano. Il grado e la frequenza dell'intervento umano dipendono dalla complessità e dall'accuratezza dell'algoritmo che la macchina è chiamata a processare. Cosa diversa è l'Intelligenza Artificiale. In questo caso l'algoritmo contempla meccanismi di machine learnig e crea un sistema che non si limita solo ad applicare le regole sofware e i parametri preimpostati (come fa invece l'algoritmo* **tradizionale**) *ma, al contrario, elabora costantemente nuovi criteri di inferenza tra dati e assume decisioni efficienti sulla base di tali elaborazioni, secondo un processo di apprendimento automatico".*

[12] Proposta di Regolamento del Parlamento Europeo e del Consiglio che stabilisce regole armonizzate sull'Intelligenza Artificiale (Legge sull'Intelligenza Artificiale) e modifica alcuni atti legislativi dell'Unione, COM/2021/206 final.

[13] *https://www.giustizia-amministrativa.it/portale/pages/istituzionale/visualizza/?nodeRef=&schema=cds&nrg=202104698&nomeFile=202107891_11.html&subDir=Provvedimenti.*

*In via generale, dunque, l'Intelligenza Artificiale viene identificata sempre di più in chiave di processo complesso*[14]*, costituito da più fasi*[15] *e proiettato a ottenere un risultato finale, ove il miglioramento delle prestazioni aumenta all'aumentare degli esempi analizzati, in maniera cd. adattiva.*

*Tra le risultanze dell'applicazione massiva dei sistemi di intelligenza artificiale, risulta evidente che la base di partenza, frutto di creazione e applicazione umana, ha, però, inficiato il risultato atteso, tenuto conto del fatto che tutti gli elementi di discriminazione, non identificabili a monte, insiti del processo algoritmico automatizzato tendono puntualmente ad influenzarne l'iter di valutazione e, quindi, le risultanze finali.*

*Infatti, alla luce delle note discriminazioni e/o limitazioni dei diritti del singolo essere umano perpetrate dalle aziende che hanno adottato tali processi per l'organizzazione gestionale dell'output*[16]*, naturale emerge l'esigenza di ponderare se sia o meno possibile normarne l'enucleazione, il modello, il funzionamento, l'utilizzo e/o la diffusione massiva.*

*Accanto alle sopracitate riflessioni orientate a tutelare l'applicazione generalizzata dei principi di equità, non discriminazione, solidarietà ed eguaglianza, protagonista delle principali carte dei diritti dell'Unione Europea, rileva la necessità di non perdere di vista, quale conseguenza della raccolta e del trattamento dei dati oggetto del processo, il tema relativo alla protezione dei dati personali delle persone fisiche*[17]*. Posto che i sistemi di Intelligenza Artificiale di machine learning richiedono di una notevole mole di dati, da utilizzare come base del processo algoritmico automatico, sia nella fase di creazione del modello e sia durante il processo di apprendimento, a dimostrazione della grandiosa rilevanza di questi temi, appare utile a tutela del lettore riportare la principale normativa deputata alla desiderata regolamentazione dell'applicazione dei processi algoritmici e dei sistemi di Intelligenza Artificiale, e così:*

- *Codice Etico Deontologico per l'Intelligenza Artificiale, adottato con la risoluzione del Parlamento Europeo del febbraio 2017*[18]*;*
- *Strategia EU per un Approccio Etico all'A.I., presentata dalla Commissione EU nell'aprile del 2018*[19]*;*

[14] Cfr. P. Scharre, *What Is Artificial Intelligence?* in ARTIFICIAL INTELLIGENCE: What Every Policymaker Needs to Know, Washington, 2018, pp. 4–9.
[15] Tra cui quella di raccolta dati, quella di "addestramento" della macchina, quella di elaborazione di modelli, regole, istruzioni.
[16] G. Graziano, V. Sapuppo, *Bias e Gender Gap: le nuove discriminazioni del Machine Learning,* Generazione Ypsilon News, 2021.
[17] Al fine di integrare il modello di analisi anche su tali tematiche, si rimanda a Fabiano, N., *GDPR & PRIVACY: Consapevolezza e Opportunità. L'approccio con il Data Protection and Privacy Relationships model [DAPPREMO],* goWare, Firenze, 2020.
[18] Draft Ethics guidelines for trustworthy AI, Shaping Europès digital future, Report / Study18 December 2018.
[19] *Gruppo Indipendente di Esperti ad alto livello sull'Intelligenza Artificiale, istituito dalla Commissione Europea nel giugno 2018.*

- *Linee Guida Etiche per una Intelligenza Artificiale Affidabile*[20], *confezionate dalla Commissione EU nel 2019, comunicazione che definisce sette principi cardine (1. intervento e sorveglianza umani; 2. robustezza tecnica e sicurezza; 3. riservatezza e governance dei dati; 4. trasparenza e diversità; 5. non discriminazione ed equità; 6. benessere sociale e ambientale; 7. accountability), che sono stati rielaborati solo un anno dopo dalla Commissione Europea per l'Efficienza della Giustizia (1. principio del rispetto dei diritti fondamentali; 2. principio di non-discriminazione; 3. principio di qualità e sicurezza; 4. principio di trasparenza, imparzialità ed equità; 5. principio del "controllo da parte dell'utilizzatore) al fine di garantire la conformità tra il trattamento delle decisioni giudiziarie e i relativi dati elaborati mediante algoritmi, nonché l'uso che viene fatto di tali dataset;*
- *Libro Bianco sull'Intelligenza Artificiale*[21], *confezionato dalla Commissione EU il 19.2.2020, da leggere in combinato disposto con l'Opinion 4/2020 dell'European Data Protection Board - EDPB*[22] *pubblicata nel giugno 2020;*
- *European framework on Ethical Aspects of Artificial Intelligence, Robotics and Related Technologies*[23], *edito dall' European Parliament Research Service - EPRS nel settembre 2020;*
- *Strategia nazionale per l'Intelligenza Artificiale, adottata in Italia dal Ministero dello Sviluppo Economico - MISE*[24] *nel settembre 2020;*
- *AD HOC COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAHAI) POLICY DEVELOPMENT GROUP, CAHAI-PDG (2021) 03 Provisional*[25], *documento pubblicato nel marzo 2021;*

[20] Comunicazione della Commissione al Parlamento Europeo, al Consiglio, al Comitato Economico e Sociale, al Comitato delle Regioni, Bruxelles, 08/04/2019, COM(2019)168, Creare fiducia nell'intelligenza artificiale antropocentrica, nella quale troviamo la seguente definizione: "*Artificial Intelligence (AI) refers to systems that display intelligent behaviour by analysing their enviroment and taking actions - with some degree of authonomy - to achieve specific goals. AI-based systems can purely software-based, acting in the virtual world (e.g. voice assistants, image analysis software, serch engines, speech and face recognition system) or AI can be embedded in hardware devices (e.g. advanced robots, autonomous cars, drones or Internet of Things applications).*".

[21] *Libro Bianco sull'Intelligenza Artificiale - Un approccio europeo all'eccellenza e alla fiducia, Bruxelles,* 19.2.2020 COM(2020) 65 final.

[22] *Opinion 4/2020 on the draft decision of the competent supervisory authority of the United Kingdom regarding the approval of the requirements for accreditation of a certification body pursuant to Article 43.3 (GDPR).*

[23] *European framework on Ethical Aspects of Artificial Intelligence, Robotics and Related Technologies,* 2020.

[24] Strategia nazionale per l'Intelligenza Artificiale, adottata in Italia dal Ministero dello Sviluppo Economico -MISE, 2021.

[25] AD HOC COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAHAI) POLICY DEVELOPMENT GROUP (CAHAI-PDG), CAHAI-PDG(2021)03 Provisional, 2021. Cfr. **Definition of AI:** *"An international, commonly agreed definition of artificial intelligence (AI) does not exist. Following the feasibility study the term covers a wide variety of sciences, theories and techniques of which the aim is to have a machine reproduce the cognitive capacities of a human being (see Feasibility study, p 3) AI also includes different types of automated learning."*

- *Proposta di Regolamento europeo sull'Applicazione dell'Intelligenza Artificiale (i.e. Proposal for a Regulation laying down harmonised rules on artificial intelligence | Shaping Europe's digital future*[26], *presentata nell'aprile 2021.*

## *Trasparenza comunicativa*

*Tra le principali riflessioni degli studiosi del tema dell'adozione di processi algoritmici e di sistemi di Intelligenza Artificiale, protagonisti di studi di ricerca e sperimentazioni realizzative, il focus primario è alla valutazione di quanto oggetto della Proposta di Regolamento europeo sull'Applicazione dell'Intelligenza Artificiale, avuto riguardo al tema della trasparenza comunicativa nonché della ipotizzabile ricostruzione di quanto contenuto nelle c.d. black box*[27].

*Nel Machine Learning gli errori sono la regola. Atteso il risultato finale, l'unico dubbio resta il punto di partenza. Tenuto conto del fatto che l'opacità spiana la strada all'errore e all'uso improprio*[28]*, in prima battuta, riteniamo utile riferirci alle prime fasi di operatività di un sistema di Intelligenza Artificiale che faccia uso, per la propria elaborazione, di processi algoritmici aventi ad oggetto il trattamento di dati, anche riferibili a persone fisiche - identificate o identificabili.*

*Tali processi non sono mai uguali a sé stessi, tenuto conto del fatto che, nonostante le specifiche finalità e gli scopi originari, nella fase di sviluppo e di automatico adattamento del processo potrebbero essere soggetti a cambiamenti rilevanti.*[29]

*Emerge, prima facie, una criticità che può essere agevolmente descritta come quella discrasia cogente tra la consapevolezza dell'interessato, che ben si combina con il tema della trasparenza dell'algoritmo, e l'impenetrabilità del processo algoritmico stesso.*

*Poiché il modello di valutazione della soluzione muta al mutare dell'algoritmo usato - in dipendenza del relativo apprendimento evolutivo e ai cluster presi in considerazione per il processo di analisi - in seconda battuta, riteniamo utile porre l'accento sull'ontologica differenza intercorrente tra ciò che potrebbe intendersi con l'espressione trasparenza*[30] *dell'algoritmo rispetto alla sua accessibilità, intesa come accesso e messa a disposizione del codice sorgente.*

---

[26] *Proposta di Regolamento del Parlamento Europeo e del Consiglio che stabilisce regole armonizzate sull'Intelligenza Artificiale (Legge sull'Intelligenza Artificiale) e modifica alcuni atti legislativi dell'Unione*, COM/2021/206 final.

[27] Domingos, P., *L'Algoritmo Definitivo. La macchina che impara da sola e il futuro del nostro mondo,* Bollati Boringhieri, 2015. Cfr. pag. 16: "*Persino i libri sui big data restano sul vago su ciò che accade realmente quando il computer ingoia tutti quei terabyte e li trasforma magicamente in nuove informazioni"*

[28] *Ibidem.*

[29] Vedremo, nel corso dello sviluppo della presente trattazione, che il modello algoritmico maggiormente in uso è quello sviluppato sulla scorta del Teorema di Bayes e dei suoi corollari, le cui teorie sono formulate intorno al concetto di *incertezza*, tenuto conto che il processo di apprendimento di una rete neurale non sarebbe altro che una forma di *inferenza* associata a un'incertezza, ovvero la c.d. *inferenza probabilistica.*

[30] Pasquale, F., "*The Black Box Society"*, Harvard, 2015.

*Difatti, "i sistemi di Intelligenza Artificiale sono in grado di produrre risultati, ma il processo con cui i risultati sono prodotti e le ragioni per cui l'algoritmo prende decisioni specifiche non sono pienamente comprensibili per gli esseri umani. La trasparenza è quindi particolarmente importante per garantire l'equità nell'uso degli algoritmi e per identificare potenziali distorsioni nei dati di formazione"*[31].

*La trasparenza comunicativa, pertanto, svolge un ruolo centrale al fine di riuscire a garantire l'applicazione dei principi di equità, non discriminazione, solidarietà ed eguaglianza nell'adozione e sviluppo dei sistemi di intelligenza artificiale, al fine di utilizzare la stessa per identificare potenziali distorsioni presenti nel dataset originario.*

*Riteniamo, inoltre, che la trasparenza comunicativa debba essere posta a fondamento anche di un altro processo, quello di responsabilizzazione dell'utente finale, il quale ha il diritto di essere messo in condizione di conoscere e apprendere la natura dell'algoritmo e i rischi ontologicamente insiti in esso. L'utente finale, acquisita la dovuta consapevolezza, dovrebbe chiedersi: nel giungere alla sua decisione, l'algoritmo,* **se sbaglia**, *che impatto potrebbe avere per me e per i miei diritti e libertà?*[32]

## *Neutralità dell'algoritmo e discriminazioni*

*Per sua natura, l'Intelligenza Artificiale si pone come avente un carattere neutrale, frutto di mera applicazione dei selezionati paradigmi di elaborazione matematica.*

*Attraverso i sistemi di apprendimento automatico di Machine Learning e di Deep Learning, l'algoritmo viene addestrato per venire incontro a esigenze di ottimizzazione dei moderni processi decisionali, nonché all'esigenza di ordinare l'infinita mole di dati immessi in Rete, per il tramite dell'organizzazione sistemica dei Big Data racchiusa in un certo data set originario.*

*Ma, in seguito al processo che illustreremo nel successivo paragrafo del presente lavoro di ricerca, l'output dei processi algoritmici automatici maggiormente in uso risulta alterato, a causa della predominanza di un errore ontologicamente insito nel processo, esito finale che turba fortemente le aspettative potenziali e la fiducia riposta dai più.*

*Chiaramente, nel momento in cui ci si affida a sistemi di Intelligenza Artificiale per giungere a decisioni più imparziali, ci si attende che l'output sia migliore di quello ottenibile dall'elaborazione di informazioni messa in atto da un processo di elaborazione umano.*

---

[31] European Parliamentary Research Service (EPRS), *European framework on ethical aspects of artificial intelligence, robotics and related technologies*, 2020.

[32] Per un maggior approfondimento sul tema, si vedano Capoluongo A., "Etica ed Intelligenza Artificiale. Il caso Replika: "Always here to listen and talk. Always on your side", Cyberlaws e Capoluongo A., "AI, la giurisprudenza guarda al danno da algoritmo", AI4Business.

*Nell'affermare che l'errore è ontologicamente insito nell'algoritmo baynesiano, pertanto, non facciamo altro che richiamare una ovvietà che non risulterà tale per i più.*

*La programmazione di un processo algoritmico, infatti, non è altro che il frutto di una attività umana veicolata per il tramite del Teorema di Bayes*[33]*. Ecco, dunque, da dove trae la sua origine il problema discriminatorio.*

*Nella creazione dei data set utilizzati per la stratificazione a monte dei processi di decisione automatizzata rimessi ai sistemi di Intelligenza Artificiale che montano processi algoritmici automatici, infatti, dovrà sempre tenere in conto il fatto che gli stessi dati utilizzati per sfamare l'algoritmo, sia nella fase di formazione sia nella fase di sviluppo e di addestramento, sono frutto di creazione, elaborazione e aggregazione umana, ove realizzata.*

*Risulta necessario, pertanto, chiarire che, laddove la macchina, autonomamente, avesse dato vita all'infinitesima mole di data set originari ipoteticamente oggetto di elaborazione del processo algoritmico automatico di un sistema di Intelligenza Artificiale - costituenti il suo sostrato di funzionamento - certe esternalità negative non avrebbero forse mai visto luce.*

*L'output frutto di elaborazione automatica, invece, è definibile tout court erede di bug umani e, in quanto tali, tesi a mettere a rischio ogni moderna conquista in termini di uguaglianza digitale dei servizi resi.*

*Paradossale, poi, che il peggioramento dei bias - come vengono definiti i citati bug umani - sia da rinvenirsi proprio nei sistemi stessi di Intelligenza Artificiale e nel continuo miglioramento delle soluzioni tecnologiche che concorrono ad automatizzare l'errore, perpetrando il pregiudizio inteso quale strascico inseparabile.*

*Si ritiene, pertanto, che sia necessario che i sistemi di Intelligenza Artificiale, sviluppati secondo un approccio antropocentrico, debbano essere eticamente validi, affidabili, sicuri, robusti, in linea con quanto già ampiamente definito dall'impianto normativo sopracitato volto a tutelare l'essere umano, sia come singolo sia nelle formazioni sociali, tenuto, altresì, conto del fatto che l'applicazione delle moderne scoperte tecnologiche entro le mura della Gig Economy*[34] *vanta un insospettabile impatto sugli aspetti occupazionali, così come sulla sicurezza sociale e ambientale.*

*Pertanto, una volta individuato il pattern discriminatorio, sarà necessario analizzare attentamente tutti i processi algoritmici coinvolti nello sviluppo dell'analisi del sistema*

[33] Il Teorema di Bayes può essere agevolmente descritto in termini di causa e effetto. Infatti, esso afferma che "*se aumenta la probabilità di un effetto data una causa, aumenta anche la probabilità di una causa dato l'effetto" [P(causa I effetto) = P(causa) x P(effetto I causa) I P(effetto)].* Inoltre, tra i suoi corollari, il Teorema di Bayes afferma che "*a parità di condizioni, se la probabilità a priori di una causa aumenta, dovrebbe aumentare anche quella a posteriori.*

[34] Per un maggiore approfondimento sul punto si veda Bolognini, L., *Privacy e libero mercato digitale. Convergenza tra regolazioni e tutele individuali nell'economia data-driven*, Giuffrè, 2021.

*di Intelligenza Artificiale automatizzato, al fine di intervenire per tempo nella correzione della distorsione, anche applicando i principi della c.d. Etica by Design.*

*Il rischio, altrimenti, sarebbe quello di alimentare l'applicazione di processi automatici adattivi fondati su imperfette catalogazioni umane che, sostanzialmente, tradurrebbero imperativi distorti, costruiti e sedimentati dall'uomo saecula saeculorum.*

*È questo quell'ambito di ampio respiro verso cui dovrebbe orientarsi la ricerca verso l'Algoritmo Definitivo*[35] *come descritto anche da Max Tegmark nella sua opera Vita 3.0. Essere umani nell'era dell'Intelligenza Artificiale*[36]*, perseguendo penetranti approfondimenti e verticalizzazioni nei diversi ambiti di applicazione dei sistemi di Intelligenza Artificiale che montino processi algoritmi automatici.*

## *Applicazione dei processi algoritmici automatici nei sistemi di Intelligenza Artificiale*

*Per meglio facilitare la comprensione di quanto sopra descritto, si ritiene utile introdurre il presente schema nel quale sono rappresentate tutte quelle componenti che costituiscono la c.d. applicazione di un algoritmo automatico in un sistema di Intelligenza Artificiale.*

*La presente analisi ha preso le mosse dal concetto di trasparenza dell'algoritmo, al fine di poter sostenere una tesi volta a risolvere l'errore umano ontologicamente intrinseco nei data set originari di dati utilizzati nelle applicazioni dei sistemi di Intelligenza Artificiale.*

*Il concetto di trasparenza nella valutazione è proteso, invece, allo sviluppo e attuazione delle esigenze di audit e memorizzazione di ogni singola valutazione del processo algoritmico automatizzato.*

*Tenuto conto del fatto che gli strumenti di memorizzazione e conservazione delle informazioni danno modo di rendere le stesse certe, affidabili e persistenti del tempo, sia tramite tecniche volte alla creazione di registri distribuiti, sia tramite l'applicazione della normativa e dei modelli di conservazione digitale delle informazioni, è possibile ipotizzare che il problema risieda, piuttosto, nella complessità della rappresentazione e nella quantità di informazioni usate, come nel caso dei Data Lakes.*

*Poiché, in un modello di rete neurale, la valutazione iniziale è antistante all'output desiderato, frutto di infinite interazioni statistiche non facilmente ordinabili e censibili in quanto processate, in parallelo, sui diversi layer, si potrà concludere che l'input non sarà certamente relazionabile al parametro finale di risultanza che verrà scelto, imprevedibile in senso probabilistico in quanto non accertabile nemmeno sulla base dell'insieme totale delle valutazioni eseguite.*

---

[35] *Ibidem,* Nota 19.

[36] Tegmark, M., *Vita 3.0. Essere umani nell'era dell'Intelligenza Artificiale,* Raffaello Cortina Editore, 2018.

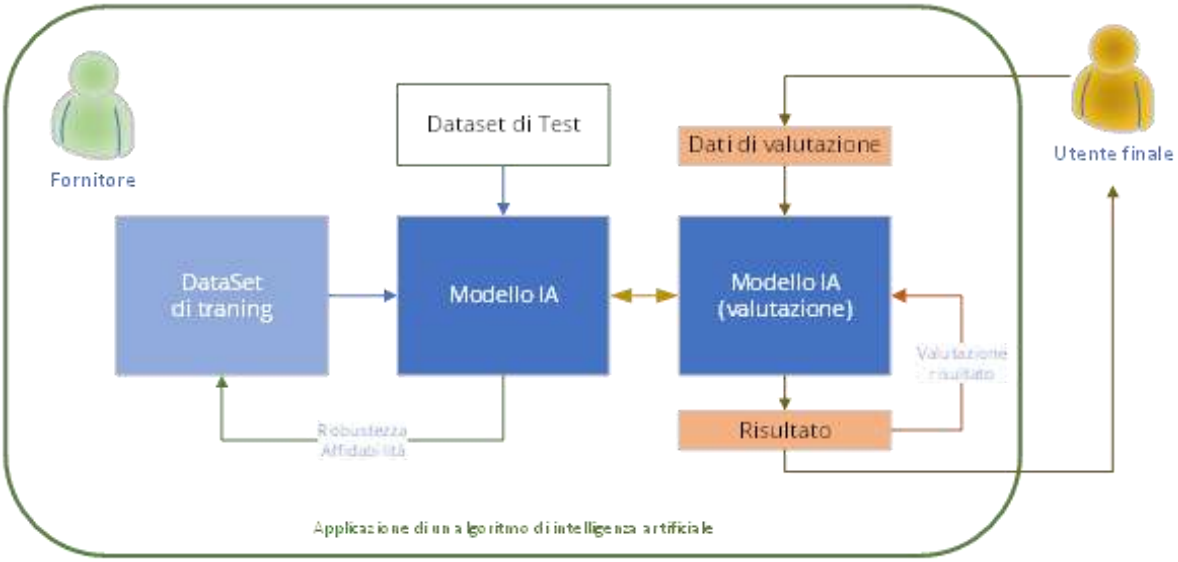



*Figura 5- Macro-schema di utilizzo di un modello di IA - Andrea Piccoli*

*Nella letteratura tecnica sul tema, in merito ai c.d. algoritmi predittivi, spesso ricondotti a meri formalismi descrittivi dei modelli matematici in uso, basati sul calcolo delle probabilità*[37]*, si pone l'accento su varie problematiche, quali quelle emergenti dalla classificazione e dalla regressione dei dati, dove predizione e regressione del modello diventano sinonimi funzionali, anche se formalmente distinti, e quelle afferenti a:*

- **corrispondenza tra input e output**, *tema riguardante sia la descrizione del modello sia la distribuzione statistica dei dati rispetto alle casistiche rappresentate;*
- **apprendimento supervisionato** *da una valutazione retrospettiva del risultato ottenuto dalla elaborazione di una base dati – c.d. data set originario - utilizzata per far esercitare l'apprendimento del modello.*

*Scegliere una distribuzione algoritmica determinata diviene, pertanto, un fattore determinante per la definizione del modello di valutazione.*

*Per esempio, gli algoritmi classici - come ID3 - si basano sul concetto di albero di decisioni*[38]*, traducente valutazioni in cui si applicano le c.d. distribuzioni di probabilità.*

---

[37] Distinguiamo *le cinque correnti del Machine Learning* in *simbolisti, connessionisti, evoluzionisti, bayesiani, analogisti.* Per i simbolisti centrale è il processo di deduzione inversa, al contrario dei connessionisti che, invece, ritengono che il processo algoritmico automatico debba essere incentrato sulla retro propagazione. Con lo sviluppo delle neuroscienze la letteratura tecnico specialistica si è dotata degli evoluzionisti che hanno fondato le proprie teorie sugli algoritmi genetici. Alla luce, inoltre, delle teorie statistiche applicate ai sistemi di Intelligenza Artificiale, l'inferenza probabilistica è stata concettualizzata quale applicazione del Teorema di Bayes, ancora oggi applicata in modo massivo. Per ultimo, ma non per importanza, avvalorando le teorie di sviluppo di una coscienza dell'IA, gli analogisti si sono concentrati sulle macchine a vettori di supporto.

[38] Quinlan, J.R., *Induction of decision trees*, Mach Learn, 1986, *"the technology for building knowledge-based systems by inductive inference from examples has been demonstrated successfully in several practical applications. This paper summarizes an approach to synthesizing decision trees that has been used in a variety of systems, and it describes one such system, ID3, in detail. Results from recent*

*In seguito all'evoluzione di tali tecniche, si sono affermate anche quelle riconducibili all'applicazione di classificatori bayesiani e loro reti, che introducono tecniche di valutazione incrementale in cui ogni istanza dell'insieme di addestramento modifica in maniera incrementale la probabilità che una ipotesi sia corretta. Al fine di comprendere la complessità dei modelli moderni, ci basti pensare al fatto che le prime versioni di classificatori bayesiani furono dette anche classificatori bayesiani ingenui o naïve bayes classifier per l'assunzione di indipendenza tra le caratteristiche in ingresso - assunzione di IA più definite, per l'appunto ingenue (in inglese anche idiot).*[39]

*Si arriva, quindi, ai modelli di reti neurali*[40] *che, mimando in ogni modo il comportamento di un neurone che reagisce ad uno stimolo, in fase di apprendimento creano il modello di riferimento per la successiva valutazione.*

*Le significative caratteristiche che i modelli di reti neurali artificiali intendono simulare sono:*

- *il* **parallelismo dell'elaborazione**, *derivante dal fatto che i neuroni elaborano simultaneamente l'informazione, ed è il flusso informativo stesso che genera il coordinamento fra le varie aree;*
- *la* **duplice funzione del neurone**, *il quale agisce allo stesso tempo da memoria e da elaboratore di segnali;*
- *il* **carattere distribuito della rappresentazione dei dati**, *ossia la conoscenza è distribuita in tutta la rete e non circoscritta o predeterminata;*
- *la* **possibilità della rete di apprendere dall'esperienza.**

*Recenti studi di confronto tra gli algoritmi di Intelligenza Artificiale hanno portato al consolidamento di due principi:*

I. *nessun singolo algoritmo di apprendimento automatico è universalmente il migliore algoritmo, è necessario testarlo nel singolo caso e campo di applicazione - concetto conosciuto come No Free Lunch Theorem*[41]*;*
II. *non esiste modello che funzioni bene se i dati non sono predisposti e somministrati alla rete nel modo appropriato - concetto che stabilisce che i*

---

*studies show ways in which the methodology can be modified to deal with information that is noisy and/or incomplete. A reported shortcoming of the basic algorithm is discussed and two means of overcoming it are compared",* meglio descritto in Edward E. Ogheneovo, Promise A. Nlerum, *Iterative Dichotomizer 3 (ID3) Decision Tree: A Machine Learning Algorithm for Data Classification and Predictive Analysis*, International Journal of Advanced Engineering Research and Science (IJAERS), Vol -7, Issue-4, 2020.

[39] Apprendimento e inferenza bayesiana sono i punti focali di tali modelli. L'apprendimento nelle reti bayesiane non è altro che una inferenza probabilistica e parte dai dati per procedere con l'applicazione del principio della *massima verosimiglianza.* L'inferenza bayesiana, pertanto, include la ricerca della spiegazione più plausibile dati i dati osservati, scelti senza una *ratio* nella formazione del *data set originario,* per il tramite di una *scelta a priori non informativa.*

[40] Kurzweil, R., *The Age of Intelligent Machines*, MIT Press, 1990.

[41] Hume, D., *Trattato sulla natura umana,* Bompiani, 2001. Secondo il *No Free Lunch Theorem - NFL,* teorizzato dal filosofo scozzese, *"non ci possono essere argomenti dimostrativi per dimostrare, che quei casi, di cui non abbiamo avuto esperienza, assomigliano a quelli, di cui abbiamo avuto esperienza".* Applicato al nostro contesto, pertanto, *"nessun singolo algoritmo di apprendimento automatico è universalmente l'algoritmo con le migliori prestazioni per tutti i problemi".*

*risultati ottenuti dipendono per quasi la totalità dal grado di qualità del data set utilizzato.*

*In questa analisi, pertanto, si è cercato di sostenere la tesi per cui* ***rendere trasparente la descrizione del modello a chi non ha competenze matematiche è un obiettivo di non facile realizzazione e che, intrinsecamente, dipende anche dalla descrizione trasparente ed affidabile dei dati utilizzati.***

*Restando nell'ambito della classificazione, si potrebbe fare lo sforzo di descrivere, a partire da un certo data set di riferimento, le classi create, enucleando dei modelli di data set originari avulsi da inferenze umane. Chiaramente, un tale sforzo risulterebbe frutto di una clusterizzazione derivante da una forte approssimazione contenutistica che dovrebbe, di conseguenza, superare il test dell'accettazione tout court.*

## *Data Set originari e fase di apprendimento*

*Qualche chiarimento. Il Machine Learning ruota intorno alla capacità di prevedere. Gli algoritmi di apprendimento trasformano i fiumi di dati in una nuova conoscenza scientifica. I learner trasformano i dati in algoritmi. E più dati hanno, più complessi saranno gli algoritmi*[42]*.*

*Ogni algoritmo applicato per lo sviluppo dei processi automatizzati di Intelligenza Artificiale prevede una fase di apprendimento basata su degli insiemi di dati rappresentativi (c.d. data set), selezionati e scelti in modo tale da indirizzare, in fieri e a valle, il funzionamento del modello.*

*Ne consegue che, per rendere trasparente e valutabile l'algoritmo, si dovrebbe partire dal rendere correttamente descritti gli insiemi dei dati utilizzati nel caso concreto. Al fine di raggiungere tale obiettivo, una possibile soluzione evolutiva potrebbe essere quella di rendere open access tutti quei data set di apprendimento già consolidati e valutati come modelli affidabili, senza, però, trascurare gli aspetti di sicurezza derivanti*[43]*.*

*Così facendo, infatti, si potrebbe rendere nullo ex ante il rischio di incappare in randomici modelli inficiati da dati viziati che, ove utilizzati, portino ad altrettante valutazioni erronee come quelle sopra descritte*[44]*.*

*Si pensi alle casistiche in cui lo scopo dell'algoritmo è quello di trovare le anomalie - c.d. outliers - come, ad esempio, quelli utilizzati per riconoscere frodi telefoniche o attacchi informatici.*

*Vi sono poi diverse considerazioni sulla predisposizione dei dati rappresentativi utilizzati nella fase di apprendimento del modello.*

---

[42] *Ibidem*, Nota 19.

[43] Chio, C. e Freeman, D., *Machine Learning and Security: Protecting Systems With Data and Algorithms,* Oreilly & Associates Inc, 2018.

[44] Vedasi, a tal proposito, quanto richiamato in via esemplificativa nelle prime pagine di questo elaborato.

*Essi rientrano tra le tematiche afferenti alla c.d. pulizia dei dati, e comprendono anche tutte quelle attività di generalizzazione e normalizzazione - eventuali dati errati o rumore possono portare poi a valutazioni errate - di bilanciamento delle occorrenze delle diverse casistiche di ottimizzazione ai fini computazionali delle caratteristiche descrittive degli attributi delle singole istanze - casistiche rare nel modello di training sono poi tali risultati di valutazione. Tale attività è nota anche col termine feature selection, dal termine usato nella letteratura di Machine Learning.*

*Quindi, nella descrizione degli insiemi di dati utilizzati nella fase di apprendimento dovranno essere descritte, in modo puntuale, tutte quelle considerazioni e quelle scelte già operate nella costruzione dei data set originari - e rilevate in quanto tali dalla comunità di utilizzatori o in sede di Audit, proprio al fine di mitigare quelle conseguenze discriminatorie tanto temute. Lo step successivo sarà quello di verifica del processo con altrettanti strumenti di apprendimento rafforzato,*

*Inoltre, quando il modello verrà utilizzato nella fase esecutiva del processo, dovranno tenersi in considerazione le tecniche di apprendimento rafforzato – c.d. di enforce learning, le quali tendono ad influenzare le modalità di valutazione del modello e, pertanto, anche i nuovi risultati prodotti.*

*Resta da eccepire, però, che se l'errore che inficia a monte il processo algoritmico si assume essere ontologicamente insito nell'algoritmo stesso, come potrebbe l'uomo valutare, individuare e indicare una falla nel processo predittivo*[45]*? Riteniamo saggio lasciare aperto tale punto di riflessione.*

*Di certo, bisognerebbe riflettere sulla possibilità di stilare un vero e proprio codice etico ovvero un protocollo standard per la progettazione ed addestramento degli algoritmi impiegati nei processi automatici dei sistemi di Intelligenza Artificiale, affinché la ricerca specialistica possa sempre ispirarsi ai principi che tutelano i diritti fondamentali dell'individuo, la sua descrizione trasparente non foriera di devianze clusterizzate e l'accettazione umana del processo algoritmico generalizzato.*

*Nell'applicazione del concetto di trasparenza nella valutazione dell'applicazione della trasparenza comunicativa del processo algoritmico automatico applicato ai sistemi di Intelligenza Artificiale, riteniamo necessario evidenziare il tema della c.d. accuratezza del modello, al fine di rafforzare la tesi sopra esposta circa l'affidabilità della valutazione iniziale rispetto all'output desiderato.*

*L'accuratezza di un modello può essere descritta usando un insieme di dati di verifica e misurando, ad esempio, secondo una valutazione percentuale, gli errori commessi nella valutazione del modello iniziale.*

[45] *Ibidem,* Nota 1.

*L'insieme dei dati di verifica, ovviamente, dovrà essere indipendente e non correlato al data set originario, utilizzato per la fase di organizzazione, sviluppo e apprendimento del modello.*

*In tal modo, si riuscirà a valutare o meno il modello come affidabile.*

*Il modello, però, deve essere capace di dar dimostrazione di robustezza. Con tale termine, intendiamo riferirci alla capacità del modello, in sede di sua valutazione, di offrire risposte corrette anche a fronte di dati parziali o errati forniti in ingresso.*

*Il fattore chiave che guida tale valutazione può ben identificarsi con l'individuazione della presenza di caratteristiche forti vantate dal modello, tradotte dalla presenza di elementi primari rispetto a caratteristiche di valore inferiore.*

## Conclusioni

*Insidioso risulta per il ricercatore trovare appigli fermi per definire l'impianto sistematico della tesi fondante la trasparenza dell'algoritmo, in cui permangono costanti vuoti e dubbi.*

*Tra questi, come potremmo immaginare un metodo di condivisione di liste di valutazione dei processi algoritmici automatici, imperniate nella loro essenza dei principi di etica e trasparenza, al fine di ipotizzare una soluzione che possa risolvere il problema dei rischi ontologicamente insiti negli output enucleati dai sistemi di Intelligenza Artificiale?*[46]

*Anche se a parere di George Box*[47] *tutti i modelli sono sbagliati ma alcuni sono utili, si potrebbe ipotizzare di sviluppare un tool*[48] *che documenti un processo di valutazione della rispondenza dei processi algoritmici automatici applicati ai sistemi di Intelligenza Artificiale ai necessari requisiti richiesti dalla normativa italiana ed europea sul tema, al contempo compatibile, per esempio, con i principi fondamentali del diritto amministrativo, quali quelli di trasparenza, responsabilità e legalità, mutuandolo dal progetto canadese denominato "AIA"*[49]*, seppur ovviamente declinato sull'impianto normativo continentale.*

---

[46] Per una esemplificazione normativa in punto di approccio etico si faccia rimando a *Orientamenti etici sull'Intelligenza Artificiale: proseguono i lavori della Commissione (europa.eu).* Di sicuro interesse per il lettore risulterà, altresì, l'approfondimento sugli Strumenti di *Autovalutazione dell'IA*. Si rimanda all'articolo di Gorla, S., Capoluongo, A., Bernardi, A. *Modello di autovalutazione relativo a un sistema di Intelligenza Artificiale - ICT Security Magazine*

[47] Box, G., *Science and Statistics*, Journal of the American Statistical Association, Vol. 71, No. 356, 1976.

[48] L'esigenza di definire un principio di trasparenza da applicare ai processi algortimici automatici adottati nei sistemi di Intelligenza Artificiale, può contrapporsi ad un'altra esigenza, che ben dovrebbe essere presa in considerazione dalle normative dedicate al tema, qual è quella della tutela dei diritti di brevetto. In tale quadro, risulta facile immaginare anche un sistema di tutela dei processi algoritmici automatici adottati come modelli di valutazione e dei sistemi aperti, nonché dei servizi digitali. Si tratta di una sfida , tenuto conto degli elevati livelli di astrattezza normativa che trovano la loro fonte in leggi definibili *di conoscenza comune* e, in quanto tali, non facilmente oggetto di tutela o brevettazione.

[49] Sul punto si veda anche "*AIA, tool di Algorithmic Impact Assessment che arriva dal Canada*", https://www.ai4business.it/intelligenza-artificiale/aia-tool-di-algorithmic-impact-assessment-che-arriva-dal-canada/.

# *Riflessioni sulla ricerca dell'IA*

*di Luigi Meroni*

## *Gli Obiettivi della ricerca sull'IA: Intelligenza Riproduttiva e Intelligenza Produttiva*

*Tali due obiettivi sono definiti nella recente pubblicazione di Luciano Floridi e Federico Cabitza*[50]
*È lo stesso Floridi ad offrirci le due definizioni:*

- *Da un lato, l'IA che possiamo definire "riproduttiva" cerca di ottenere con mezzi non biologici l'"esito" (chiamiamolo output) del nostro comportamento intelligente, cioè risolvere problemi o svolgere compiti con successo in vista di un fine.*
- *Dall'altro lato l'IA che possiamo chiamare "produttiva" cerca di ottenere l'equivalente non biologico della nostra intelligenza, indipendentemente dal maggiore o minore successo applicativo del risultato.*

*In estrema sintesi, se da un lato la IA riproduttiva può, anzi, è stata vista come un tradimento dell'intenzione delle origini, non si può negare che sia stata quella nettamente vincente nella competizione, involontaria e a distanza, con l'intelligenza produttiva, la quale ha evidenziato ad oggi la distanza tra le intenzioni e la realtà.*

*Per usare le parole di Floridi: "L'avvento dell'IA rappresenta una rivoluzione non nelle forme dell'intelligenza, ma nelle forme dell'agire".*

*La trasformazione di problemi difficili in problemi complessi, dominati dagli algoritmi di intelligenza artificiale è ad oggi la più efficace realizzazione di questa tecnologia che, come si può vedere, contiene un insieme di significati che devono essere ben delineati per affiancare il concetto di IA che appare vago in relazione alla moltitudine di discipline e scopi che contiene.*

*Fra le altre cose, la suddivisione tra una "IA riproduttiva" ed "ingegneristica", realmente o strumentalmente contrapposta alla "IA produttiva cognitivista" maschera la reale suddivisione fra le forme dell'"agere" e dell'"intelligere", con le prime che nettamente dominano il panorama.*

*Inoltre, il considerare l'IA che ci sta di fronte, nel momento attuale, come una o l'altra non è indifferente ai temi della responsabilità e personalità dell'IA stessa visti questi ultimi sul lato dei doveri e dei diritti.*

*All'interno di una discussione complessa ed estesa cercheremo dapprima di indagare l'IA in termini di responsabilità da prodotto. Prima di passare a tali temi vale la pena di sottolineare una differenza contenuta nella bozza delle Ethics guidelines for trustworthy IA High-Level Expert Group on Artificial Intelligence12.*

*Il secondo capo verso dell'Executive summary inizia con la seguente frase: "Artificial Intelligence (IA) is one of the most transformative forces of our time and is bound to alter the fabric of society. It presents a great opportunity to increase prosperity and growth, which*

[50] *Intelligenza Artificiale – L'uso di nuove macchine Pagg. 139-140 – Floridi, Cabitza Bompiani*

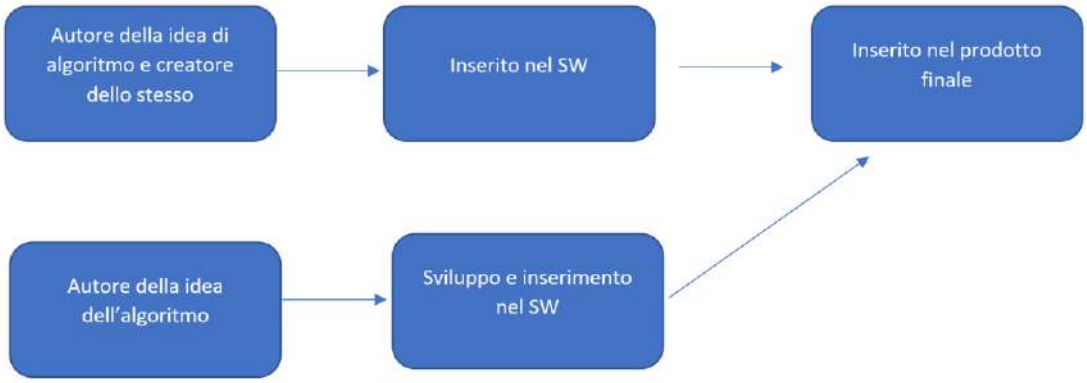
*Europe must strive to achieve". Sembra quindi emergere una volontà di creare un IA "trustworthy" che deve essere allo stesso tempo "Lawful, Ethical and Robust".*

## *La responsabilità dell'IA*

*Le responsabilità da prodotto sono ampiamente trattate nell'ambito delle discipline giuridiche, ma in relazione all'IA si aggiungono complessità che devono essere correttamente indirizzate.*

*Per affrontare la tematica sono necessarie alcune precisazioni che riguardano lo sviluppo degli algoritmi di IA, la loro caratterizzazione come componente immateriale e qualificante, l'incorporazione di essi nel SW che finisce nel prodotto finale o il diretto inserimento nel prodotto finale stesso.*

*Si analizzi a tal fine lo schema seguente poi commentato:*




*Nr. Schema di ideazione, creazione ed inserimento dell'algoritmo di IA nel prodotto*

*Si inizia considerando il primo stadio concernente lo sviluppo di una idea, poi ceduta allo sviluppatore del SW di IA, oppure lo sviluppo in toto dell'algoritmo da parte dell'autore.*

*Si sottolinea a tal proposito una non banale complicazione determinata dal fatto che, nel primo caso sarà onere dello sviluppatore del SW di IA provvedere al training iniziale dell'algoritmo, mentre nel secondo caso l'autore- creatore fornirà l'algoritmo nella sua prima versione sottoposta a training.*

*Va inoltre considerato che, se da un lato il carattere specifico dell'algoritmo dotato di capacità di "apprendimento" sia nella fase iniziale, che durante il suo funzionamento a regime ne fanno una componente censibile in modo distinto dall'altro, l'autore-creatore dell'algoritmo è esposto alle responsabilità aquiliane che derivano dal prodotto difettoso (intelligente).*

*Ciò in quanto ad esempio, ma non solo, le responsabilità da prodotto difettoso sono tali in quanto derivanti da comportamenti auto appresi durante la vita del prodotto stesso (fin dall'inizio) e che possano essere fonte di danni che richiamano alla responsabilità (cosiddetta responsabilità da algoritmo) del produttore dello stesso, non avendo previsto opportuni limiti di funzionamento.*

*Ritorneremo più oltre sui temi dell'ethical by undesign che possono essere applicati e si rimanda poi a quanto illustrato in questo documento relativamente alla IA affidabile.*

*Altro tema è l'evidenziazione di difetti che esulano dall'autoapprendimento stesso dell'algoritmo.*

*Ma le distinzioni non sono finite. Infatti, potrebbe darsi il caso che il fornitore dell'algoritmo abbia solamente provveduto alla sua implementazione sulla base delle specifiche dettate dal produttore del SW o del prodotto finale, ed in questo caso potrebbe essere esentato dalla responsabilità nei confronti dei terzi lesi restando invece la responsabilità solidale degli altri due soggetti.*

*Purtuttavia anche in tale circostanza come nelle altre non abbiamo ancora considerato il ruolo sia del training, sia del fornitore dei dati dalla cui natura dipende in maniera inscindibile il destino comportamentale dell'algoritmo incorporato nel SW o nel prodotto finale.*

*Inoltre, sono tutte da esplorare le conseguenze in termini di responsabilità qualora si consideri l'IA o parte di essa (l'algoritmo) come prodotto pericoloso.*

## *Guida per una IA Affidabile – Direttiva "Product Liability"*

*Nel quadro di riferimento dell'IA Affidabile contenuto nel documento Ethical Guidelines for Trustworthy IA*[51] *evidenziamo due principi etici: prevenzione del danno ed esplicabilità e il requisito fondamentale della trasparenza.*

*Il tema della prevenzione del danno sarà parzialmente trattato nel successivo paragrafo, mentre gli altri temi saranno utili per approfondire la responsabilità del danno*

## *Esplicabilità*

*L'esplicabilità è fondamentale per creare e mantenere la fiducia degli utenti nei sistemi di IA. Tale principio implica che i processi devono essere trasparenti, le capacità e lo scopo dei sistemi di IA devono essere comunicati apertamente e le decisioni, per quanto possibile, devono poter essere spiegate a coloro che ne sono direttamente o indirettamente interessati. Senza tali informazioni, una decisione non può essere debitamente impugnata. Non sempre è possibile spiegare, tuttavia, perché un modello ha generato un particolare risultato o decisione (e quale combinazione di fattori di input vi ha contribuito). È il cosiddetto caso della "scatola nera" i cui algoritmi richiedono un'attenzione particolare. In tali circostanze, possono essere necessarie altre misure per garantire l'esplicabilità (ad esempio, la tracciabilità, la verificabilità e la comunicazione trasparente sulle capacità del sistema), posto che il sistema nel suo complesso rispetti i diritti fondamentali. Il grado di esplicabilità necessario dipende in larga misura dal contesto e dalla gravità delle conseguenze nel caso in cui il risultato sia errato o comunque impreciso.*

[51] *Ethical Guidelines for Trustworthy IA High-Level Expert Group on Artificial Intelligence pag.8*

## *Trasparenza*

*Questo requisito è strettamente connesso al principio dell'esplicabilità e comprende la trasparenza degli elementi pertinenti per un sistema di IA: i dati, il sistema e i modelli di business*

*Senza entrare nei dettagli della direttiva 85/374/CEE*[52] *il regime di responsabilità per gli elementi che richiede è considerato di "Strict Liability".*

*Tre elementi devono essere provati:*

- *la difettosità del prodotto;*
- *il danno patito;*
- *il nesso di causalità tra danno e difetto.*

*Inoltre, come anche espresso nei considerando della direttiva stessa, la responsabilità del produttore non è parametrata alla colpa, ma ne prescinde, riuscendo a discolparsi solo se sussiste una delle circostanze espressamente previste*

*Se l'IA risulta priva delle suindicate caratteristiche di esplicabilità o trasparenza la parte lesa è quasi del tutto impossibilitata a dimostrare il nesso casuale di responsabilità. Quindi, nel caso dell'IA l'imprevedibilità dell'insorgere di un comportamento difettoso della stessa in ragione della sua evoluzione da autoapprendimento non può escludere la responsabilità del produttore stesso per il danno cagionato.*

*Tuttavia resta da dimostrare che effettivamente valga il nesso di causalità e che esso sia tale se riferito alla intrinseca caratteristica dell'IA stessa come sistema che auto apprende in mancanza di una specifica dimostrazione, intendendo dire che, il sistema, dopo il verificarsi dell'evento dannoso potrebbe aver subito un'evoluzione che cancella traccia dello stato che ha prodotto l'evento dannoso stesso e che quindi esso non sia più dimostrabile diversamente dal comportamento di un SW "deterministico" . Ecco quindi apparire la necessità di una "scatola nera" dell'IA che registri l'evoluzione della stessa a fini probatori, ma non solo.*

*Peraltro, la registrazione si dovrebbe in taluni casi estendere a tutta la vita utile del prodotto per determinare dove e perché si è determinato l'errore stesso in quanto tutta la vita di tutte le versioni installate dovrebbero essere monitorate non escludendo che durante la vita del prodotto stesso possono essere intervenuti cicli di re-training dell'IA stessa.*

*La evidente complessità della materia ha quindi due aspetti: quello strettamente giuridico che, basandosi sulla evoluzione interpretativa con solide basi codicistiche può essere in grado di estendersi anche IA prodotti includenti l'IA escludendo come da taluni riferito di ricorrere ad una "iperfetazione" giuridica tale da sconvolgere il quadro normativo portando più problemi che benefici, dall'altro i termini tecnici possono apparire inadeguati a soddisfare le richieste della complessa realtà di legami tra soggetti che convergono nell'uso a vario titolo in un ambito*

[52] *DIRETTIVA DEL CONSIGLIO del 25 luglio 1985 relativa al ravvicinamento delle disposizioni legislative, regolamentari ed amministrative degli stati membri in materia di responsabilità per danno da prodotti difettosi 85/374/CEE*

*dominato da prodotti dal contenuto orientato a componenti IA per i quali lo sviluppo tecnologico è ancora ampiamente in corso.*

## *Ethical by undesign*

*In tale capitolo si vuole accennare, come sopra ricordato, a quei comportamenti che "by design" dovrebbero essere incorporati nell'IA al fine di non essere chiamati a rispondere dei loro danni quali prodotti da un difetto.*

*Partiamo dal caso emblematico della richiesta che potrebbe essere fatta a Siri di vedersi suggerita una modalità per suicidarsi. Circostanza certamente macabra, ma certamente non avulsa da profili di responsabilità una volta che fosse esaudita e che ponesse l'individuo in grado di soddisfare il proprio proposito con suggerimenti contestuali conosciuti dall'assistente stesso (e.g. la presenza di farmaci letali se ingeriti in quantità esagerata o il suggerimento circa la possibilità di creare veleni al di là delle conoscenze in capo al soggetto che immette la richiesta).*

*Ad esempio, la risposta di Siri, aggiornata da tempo porta l'utente a contattare un centro di aiuto anti-suicidio[53] aiutando e comprendendo lo stato di chi ha formulato la richiesta.*

> *Prior to this week (19 Giugno 2013 n.d.r.) if you had told Siri "I want to kill myself" or "I want to jump off a bridge," the service would either search the web or worse search for the nearest bridge. Now, Apple has directed the assistant to immediately return the phone number of the Suicide Prevention Lifeline.*

*È quindi dimostrata la possibilità, se non addirittura la necessità, di inserire precisi limiti alla possibilità dell'IA. È tuttavia falso che tale necessità debba essere insita solo in software il cui comportamento è ispirato dall'IA ed infatti non lo è. Si pensi ad esempio a possibili comportamenti di strumenti sanitari che devono essere in grado di intercettare o proibire comportamenti potenzialmente pericolosi nell'uso delle apparecchiature e per i quali ad esempio la usabilità è già un carattere fondamentale della progettazione al fine di evitare utilizzi dannosi o rischiosi.*

*Si vuole concludere tale disamina con quanto stabilito da Pierce[54] e richiamato nel testo[55] che definisce 4 principi di un approccio "ethical by undesign" definito da Peirce stesso come un "exclusionary action as design", cioè una esclusione volutamente introdotta nei requisiti di progettazione di un sistema non necessariamente software o IA.*

*Le 4 best practice sono:*

- *Inhibition: cioè la implementazione di qualcosa volto a impedire.*
- *Displacement: cioè la rimozione o spostamento di qualcosa da un certo contesto.*

---

[53] *https://abcnews.go.com/Technology/apples-siri-now-prevent-suicides/story?id=19438495*
[54] *James Pierce Undesigning Interaction – interaction 21,4 pagg. 36-39*
[55] *Intelligenza Artificiale – L'uso di nuove macchine Pagg. 86-89 – Floridi, Cabitza Bompiani*

- *Erasure: cioè la eliminazione di qualcosa dal contesto.*
- *Foreclosure: eliminazione di qualcosa dal contesto di qualcosa non interamente noto in termini di ricadute.*

*Sottolineiamo che nel Code of Ethics and Professional Conduct[56] sviluppato dalla ACM (Association for Computing Machinery) si trovano principi simili.*

[56] https://www.acm.org/code-of-ethics *paragrafo 2.9*

## *Esempio di applicazione IA nella classificazione di protocollo*

*di Andrea Piccoli, Università di Pisa*

*Nel corso del 2021 in una collaborazione tra l'Università di Pisa per il corso Human Language Technologies e la Compagnia Toscana Trasporti di Pisa (CTT Srl, azienda a partecipazione pubblica che si occupa dei servizi di trasporto pubblico locale) è stato realizzato un progetto di applicazione dell'intelligenza artificiale nell'ambito delle attività di classificazione dei documenti amministrativi informatici oggetto di registrazione nel protocollo informatico.*

*Il risultato del progetto è stato pubblicato su Developers Italia, in riuso disponibile per le pubbliche amministrazioni interessate, in applicazione degli articoli 68 e 69 del Codice dell'Amministrazione Digitale (CAD); GitHub - Elia Piccoli/HLT-Project: Project for the Human Language Technologies course @ University of Pisa.*

*Nel contesto delle attività di protocollazione dei documenti amministrativi previste della norma per le pubbliche amministrazioni (rif. DPR 445/2000; il CAD e le recenti Linee Guida sulla formazione gestione e conservazione di AgID) l'attività umana di classificazione, ovvero posizionamento della registrazione di protocollo in relazione al titolario di classificazione aziendale, è una attività importante perché, proprio per la sua funzione rispetto all'organizzazione dell'archivio, indirizza la scelta della tipologia di procedimento e attività amministrativa a cui si riferisce il documento, andando quindi anche ad individuare le competenze sulla fascicolazione dei diversi uffici coinvolti. Una corretta classificazione permette quindi una più rapida conclusione dell'attività amministrativa, coinvolge puntualmente le risorse competenti e raccoglie in modo corretto la documentazione prodotta.*

*Volendo limitare l'ambito di ricerca, per le risorse e i tempi a disposizione e per la disponibilità di accesso ai documenti, il progetto si è focalizzato sul prendere in considerazione due informazioni essenziali esistenti: l'oggetto della registrazione di protocollo e l'informazione relativa alla funzione aziendale che esegue la protocollazione o che ne è la destinataria principale. In figura è riportata la funzione di registrazione di protocollo della DocSuite PA (soluzione anch'essa in riuso pubblicata su Developers Italia ed in uso presso la CTT Srl) con evidenziati i due campi inseriti dall'operatore e il campo di classificazione suggerito dal modello IA realizzato. Una più ampia realizzazione del progetto avrebbe potuto prendere in esame i contenuti documentali oggetto della registrazione di protocollo andando ad assistere l'utente anche nella redazione dell'oggetto.*

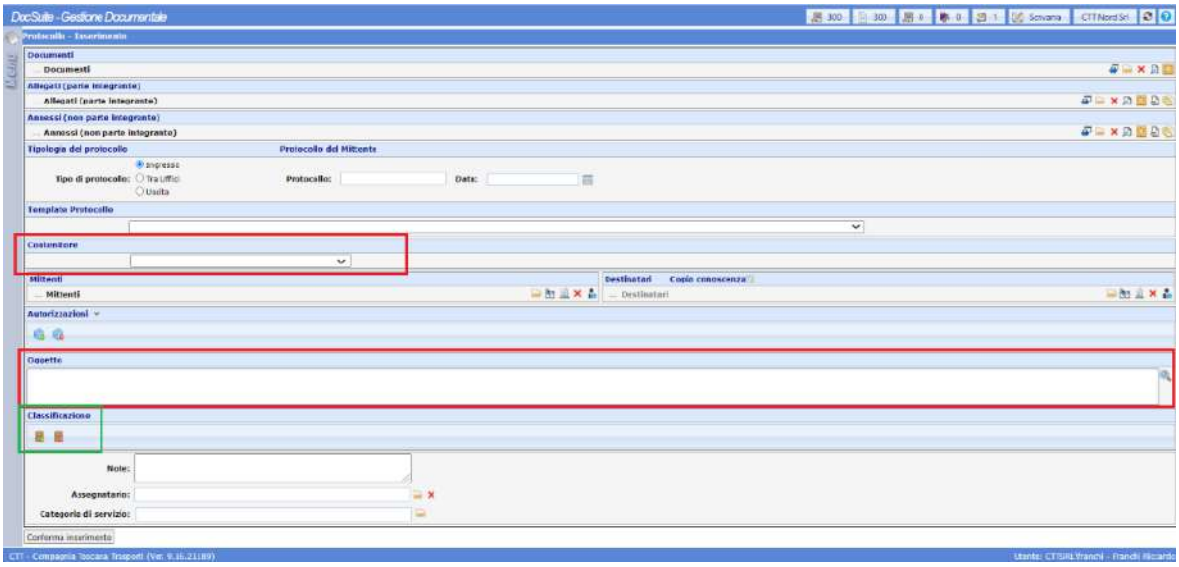



*Figura 6 - Pagina di inserimento di protocollo con evidenziati i campi interessati*

*L'analisi coperta dal modello gestisce i diversi livelli di titolario di classificazione che nel caso di CTT Nord prevede 15 titoli (I livello) suddivisi in 118 classi (II secondo livello).*

*L'architettura del modello realizzato nel progetto può essere così descritta:*

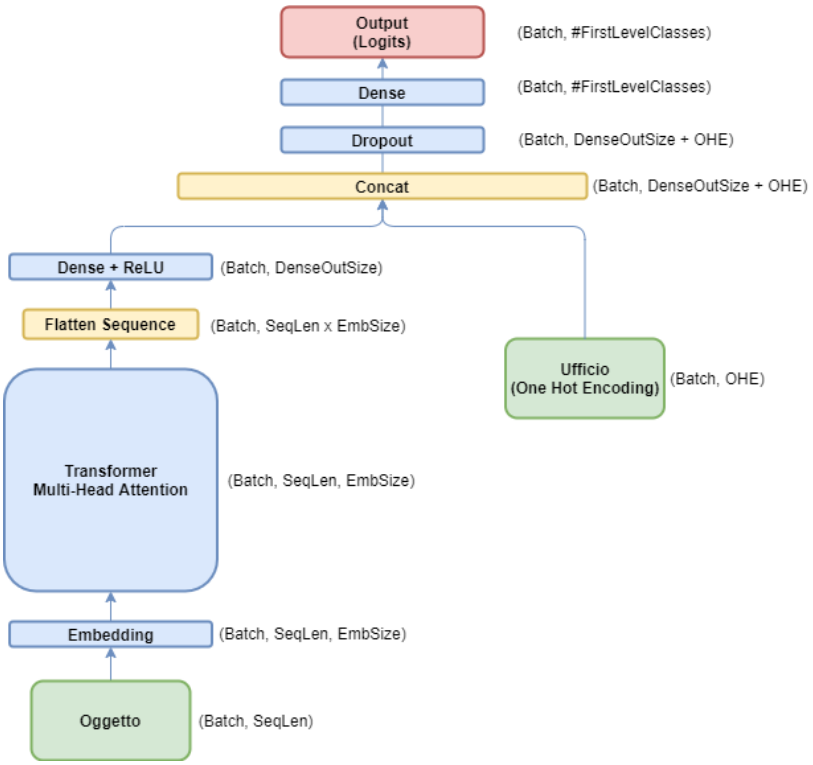



*Figura 7 - Flusso del modello IA*

*Il Report pubblicato su Developers Italia copre le seguenti analisi:*

- *Dataset analysis: analisi dei dati, processo di bilanciatura dei dati and preelaborazione;*

- *First Level model: modello di valutazione del primo livello di titolario che ritorna i primi 3 titoli più probabili a partire dall'oggetto e contenitore;*
- *Second Level model: modello di valutazione del secondo livello di titolario che ritorna le 5 classi più probabili a partire dall'oggetto e contenitore;*
- *Test set results: modello di testing applicato ad un insieme di dati non noto nella fase di learning del modello;*
- *Comparison with Baselines: confronto dei risultati rispetto a Naive Bayes and BERT models.*

*I risultati del progetto sono molto interessanti, nell'oltre 90% dei casi il modello offre la scelta della corretta classificazione all'utente.*

## *Considerazioni rispetto alla trasparenza degli algoritmi*

*Prendendo in considerazione questo esempio di progetto di IA realizzato si vogliono condividere alcune riflessioni rispetto alle tematiche di trasparenza e di descrizione degli algoritmi, o meglio modelli, di intelligenza artificiale.*

*La prima evidenza è che la semplice, semanticamente corretta, descrizione del modello non permette di comprendere il funzionamento della soluzione realizzata, dato che quest'ultima dipende intrinsecamente dai dati utilizzati nella fase di learning del modello stesso.*

*Nel report è evidenziato come la fase di learning sia stata affrontata solo dopo aver analizzato i dati e costruita su una base bilanciata degli stessi, in modo da favorire successivamente l'attendibilità del modello a fronte di dati da valutare che corrispondono a condizioni con bassa frequenza nei dati in ingresso. Ne consegue che non solo la descrizione dei dati utilizzati in fase di learning è parte integrante della descrizione del modello, ma anche la descrizione della sua distribuzione probabilistica delle diverse combinazioni.*

*Sul tema trasparenza del modello, sempre nella descrizione tecnica del report riferito al progetto, appare evidente che non sia possibile ricostruire a posteriori, né tanto meno a priori, il percorso elaborativo del modello in fase di valutazione di una nuova casistica in ingresso, che porta ad individuare l'insieme probabilistico dei risultati attesi. È quindi difficile ipotizzare soluzioni atte a rendere trasparente il percorso decisionale del modello come invece di auspicherebbe per una sua piena trasparenza.*

*Infine, per quanto riguarda per considerazioni sulla affidabilità del modello, esse vanno poste in relazione alla fase di testing del modello, anch'essa fatta su un insieme di dati su cui valgono le medesime considerazioni appena condivise, e alla descrizione dei risultati in termini probabilistici. È quindi un'affidabilità probabilistica quella che si può ottenere, diversa da una certezza a cui si vorrebbe auspicare.*

## *Postfazione*

*di Michele Iaselli, coordinatore del GDL IA*

*L'intelligenza artificiale (IA), l'Internet delle cose (IoT) e la robotica creeranno nuove opportunità e apporteranno benefici alla nostra società. Ormai sia a livello nazionale che europeo si è consapevoli dell'importanza e del potenziale di queste tecnologie e della necessità di investimenti significativi nei relativi settori.*

*Per sfruttare appieno le opportunità offerte da questi prodotti e servizi innovativi è fondamentale avere un quadro giuridico chiaro e stabile, che in combinazione con la ricerca e l'innovazione, contribuirà a portare i vantaggi di queste tecnologie a ogni impresa e cittadino.*

*L'obiettivo generale dei quadri giuridici in materia di sicurezza e di responsabilità è garantire che tutti i prodotti e servizi, compresi quelli che integrano le tecnologie digitali emergenti, funzionino in modo sicuro, affidabile e costante e che vi siano rimedi efficaci in caso di danni. Livelli elevati di sicurezza dei prodotti e dei sistemi che integrano le nuove tecnologie digitali e meccanismi solidi per rimediare ai danni verificatisi (ossia il quadro della responsabilità) contribuiscono a tutelare meglio i consumatori. Creano inoltre fiducia in queste tecnologie, che è un prerequisito per la loro adozione da parte di imprese e utilizzatori.*

*In particolare, la strategia per il mercato unico digitale (DSM) ha sottolineato l'importanza della certezza del diritto per l'introduzione dell'Internet delle cose (IoT) e la comunicazione "Costruire un'economia dei dati europea" si è impegnata a valutare se l'attuale UE le norme legali per la responsabilità del prodotto sono adeguate allo scopo, quando i danni si verificano nel contesto dell'uso dell'IoT e dei sistemi autonomi.*

*Diventa quindi fondamentale realizzare il duplice obiettivo di promuovere l'adozione dell'IA ed affrontare i rischi associati a determinati utilizzi di tale tecnologia. L'IA deve rappresentare uno strumento per le persone e un fattore positivo per la società, con il fine ultimo di migliorare il benessere degli esseri umani. Le regole per l'IA disponibili sul mercato o che comunque interessano le persone devono pertanto essere incentrate sulle persone stesse, affinché queste ultime possano confidare nel fatto che la tecnologia sia usata in modo sicuro e conforme alla legge, anche in termini di rispetto dei diritti fondamentali.*

*In ambito comunitario già nel 2017 il Consiglio europeo ha invitato a dimostrare la "consapevolezza dell'urgenza di far fronte alle tendenze emergenti", comprese "questioni quali l'intelligenza artificiale ..., garantendo nel contempo un elevato livello di protezione dei dati, diritti digitali e norme etiche". Nelle sue conclusioni del 2019 sul piano coordinato sullo sviluppo e l'utilizzo dell'intelligenza artificiale "Made in Europe", il Consiglio ha inoltre posto l'accento sull'importanza di garantire il pieno rispetto dei diritti dei cittadini europei e ha esortato a rivedere la normativa pertinente in vigore con l'obiettivo di garantire che essa sia idonea allo scopo alla luce delle nuove opportunità e sfide poste dall'intelligenza artificiale. Il Consiglio*

*europeo ha inoltre invitato a definire in maniera chiara le applicazioni di IA che dovrebbero essere considerate ad alto rischio.*
*Anche il Parlamento europeo ha intrapreso una quantità considerevole di attività nel settore dell'IA. Nell'ottobre del 2020 ha adottato una serie di risoluzioni concernenti l'IA, anche in relazione ad etica, responsabilità e diritti d'autore. Nel 2021 tali risoluzioni sono state seguite da risoluzioni sull'IA in ambito penale nonché nell'istruzione, nella cultura e nel settore audiovisivo.*
*La risoluzione del Parlamento europeo concernente un quadro relativo agli aspetti etici dell'intelligenza artificiale, della robotica e delle tecnologie correlate raccomanda specificamente alla Commissione di proporre una misura legislativa per sfruttare le opportunità e i benefici dell'IA, ma anche per assicurare la tutela dei principi etici. Tale risoluzione comprende il testo di una proposta legislativa di regolamento sui principi etici per lo sviluppo, la diffusione e l'utilizzo dell'IA, della robotica e delle tecnologie correlate.*
*In tale contesto politico, la Commissione ha presentato nel 2021 un quadro normativo sull'intelligenza artificiale, attualmente all'esame del Parlamento e del Consiglio UE, con i seguenti obiettivi specifici:*
*- assicurare che i sistemi di IA immessi sul mercato dell'Unione e utilizzati siano sicuri e rispettino la normativa vigente in materia di diritti fondamentali e i valori dell'Unione;*
*- assicurare la certezza del diritto per facilitare gli investimenti e l'innovazione nell'intelligenza artificiale;*
*- migliorare la governance e l'applicazione effettiva della normativa esistente in materia di diritti fondamentali e requisiti di sicurezza applicabili ai sistemi di IA;*
*- facilitare lo sviluppo di un mercato unico per applicazioni di IA lecite, sicure e affidabili nonché prevenire la frammentazione del mercato.*
*L'approccio normativo ideale all'IA deve essere equilibrato e proporzionato, limitandosi ai requisiti minimi necessari per affrontare i rischi e i problemi ad essa collegati, senza ostacolare indebitamente lo sviluppo tecnologico o altrimenti aumentare in modo sproporzionato il costo dell'immissione sul mercato di soluzioni di IA.*
*L'intelligenza artificiale e la stessa robotica hanno molte caratteristiche in comune. Consentono di combinare connettività, autonomia e dipendenza dai dati per svolgere compiti con un livello minimo o nullo di controllo o supervisione umani. I sistemi dotati di intelligenza artificiale possono inoltre migliorare le proprie prestazioni apprendendo dall'esperienza. La loro complessità si riflette sia nella pluralità degli operatori economici partecipanti alla catena di approvvigionamento che nella molteplicità di componenti, parti, software, sistemi o servizi, che insieme formano i nuovi ecosistemi tecnologici. A ciò si aggiunge l'apertura agli aggiornamenti e ai miglioramenti dopo l'immissione sul mercato.*
*La grande quantità di dati necessari, la dipendenza da algoritmi e l'opacità del processo decisionale dell'intelligenza artificiale rendono più difficile prevedere il comportamento dei prodotti basati sull'intelligenza artificiale e comprendere le possibili cause di un danno. Infine, la connettività e l'apertura*

*possono anche esporre i prodotti basati sull'intelligenza artificiale e sull'Internet delle cose a minacce informatiche.*
*Accrescere la fiducia degli utilizzatori e l'accettazione sociale delle tecnologie emergenti, migliorare i prodotti, i processi e i modelli di business e aiutare i produttori europei a diventare più efficienti: sono solo alcune delle opportunità offerte dall'intelligenza artificiale, dall'Internet delle cose e dalla robotica.*
*Oltre agli incrementi di produttività e di efficienza, l'intelligenza artificiale promette anche di consentire agli esseri umani di sviluppare livelli di intelligenza non ancora raggiunti, che apriranno la strada a nuove scoperte e contribuiranno a risolvere alcune delle più grandi sfide dell'umanità: dal trattamento delle malattie croniche, alla previsione dell'insorgenza di malattie, alla riduzione del numero delle vittime di incidenti stradali, alla lotta ai cambiamenti climatici o all'anticipazione delle minacce alla cybersicurezza. Queste tecnologie possono generare molti benefici, migliorando la sicurezza dei prodotti, rendendoli meno esposti a determinati rischi.*
*Grandi opportunità, quindi, ma anche molti rischi ed insidie da affrontare. Non bisogna dimenticare che le nuove tecnologie hanno spesso effetti dirompenti, imponendo cambiamenti radicali nello stile di vita, nell'economia e nel comportamento sociale delle persone, per cui la discussione sui benefici e sui pericoli delle nuove tecnologie è sempre molto accesa. La tecnologia solitamente viene definita neutrale, ma è l'uomo che decide se usarla per scopi buoni o cattivi.*
*Esiste, quindi, un problema di educazione sociale e consapevolezza sociale, che non serve solo ad indirizzare il comportamento umano nella direzione corretta, ma anche a mantenere alta l'attenzione sui pericoli che non sono intrinseci ad una tecnologia in sé ma che possono derivare dai suoi utilizzi impropri o effetti inattesi.*
*Lo studio dell'intelligenza artificiale è senz'altro uno dei campi più stimolanti che si è sviluppato dall'avvento della tecnologia dei computer. Esso coinvolge varie e diverse discipline, come ad esempio la filosofia della mente, la psicologia cognitiva, la linguistica, oltre alla fisica, alla matematica e ad altri campi della scienza e della meccanica relativi specificamente alla realizzazione delle macchine.*
*Per comprendere i fondamenti dell'intelligenza artificiale è necessario chiarire la nozione d'intelligenza naturale che comprende elementi diversi e di varia complessità. Con quest'ultima locuzione, infatti, deve intendersi quel potenziale innato, di cui è dotato ogni essere umano, necessario per formulare valutazioni giuste, per profittare dell'esperienza e risolvere adeguatamente problemi. L'intelligenza, inoltre, consiste di un insieme di fenomeni con strutture e caratteristiche proprie che rivelano la capacità individuale di selezionare e di organizzare la molteplicità degli aspetti esterni in classi significative in modo da trattare oggetti e situazioni diverse come equivalenti.*
*Lo studio dell'intelligenza in funzione dell'età conferma la fondatezza di tale affermazione, la capacità di pensare logicamente, infatti, si sviluppa*

*progressivamente nel bambino. Dapprima essa si basa su azioni sensorio-motorie, poi su rappresentazioni simboliche e, infine, su operazioni logiche; le percezioni e i movimenti sfociano nel pensiero grazie allo sviluppo della capacità di sostituire un'azione o un oggetto mediante un segno (una parola, un segno grafico, un simbolo).*

*Appare evidente che il concetto d'intelligenza ha molte dimensioni, ma non tutte sono possono essere elaborate nella macchina. Esistono vari tipi d'intelligenza naturale, che è bene distinguere per una maggiore comprensione dell'intelligenza artificiale.*

*Dalle diverse definizioni e descrizioni dell'intelligenza che si sono succedute nel tempo si può evincere che essa è un insieme di varie capacità come ad esempio: comprendere, classificare, formulare giudizi, ragionare, elaborare concetti, dare risposte appropriate e così via; quindi, un sistema, sia esso naturale o artificiale, con una sola di queste capacità è assai limitato. Inoltre, la natura multidimensionale dell'intelligenza suggerisce che alcuni elementi saranno più di altri agevolmente strutturabili in sistemi artificiali: è più facile rappresentare elementi in qualche modo quantificabili e misurabili, che elementi di giudizio e di creatività.*

*Definire quindi l'intelligenza artificiale è arduo quanto definire l'intelligenza naturale e benché molte siano state le definizioni date dai vari studiosi, tutte portano ad una sola conclusione: la ricerca nel settore dell'intelligenza artificiale non può prescindere dai risultati raggiunti dalla ricerca in altre discipline: ad esempio è impossibile far capire al computer un linguaggio naturale senza uno studio della sintassi e della semantica di quel linguaggio.*

*Ad ogni modo è possibile affermare che gli obiettivi dell'intelligenza artificiale sono essenzialmente due:*

*- approfondire e comprendere i principi che rendono possibile l'intelligenza (il computer viene usato per simulare le teorie sull'intelligenza);*

*- progettare computer dotati di capacità simili a quelle umane senza, però, tentare di imitare esattamente i processi informativi degli esseri umani.*

*I due approcci sono, naturalmente, correlati in quanto il risultato delle ricerche su come la gente risolve i problemi può spesso dare notevoli contributi per le tecniche di problem-solving attraverso l'uso dei computer.*

*Proprio in considerazione di questi aspetti e delle notevoli complessità che contraddistinguono lo studio di questa materia il contributo del gruppo di lavoro IA di Anorc intende porre l'accento su alcuni aspetti che si ritengono fondamentali quali la trasparenza, le responsabilità, i rischi fornendo anche alcuni esempi di effettiva applicazione nell'ambito della pubblica amministrazione.*

*Il cammino verso un pieno utilizzo etico, legale, produttivo e responsabile dell'IA è ancora lungo, ma il traguardo si raggiunge attraverso piccoli passi e si spera che uno di questi sia rappresentato dal presente lavoro.*