Intelligenza artificiale e Giustizia amministrativa: strategie di impiego, metodologie e sicurezza

Sommario: 1. Introduzione. 2. L'IA nella Giustizia amministrativa: applicazioni consolidate e progetti PNRR. 3. Precondizioni dell'avvio dei progetti di impiego delle tecnologie IA e sviluppi prefigurabili. 4. Linee di indirizzo seguite e ricadute pratiche. 5. Casi d'uso e ruolo del giudice nella definizione dell'architettura dei prodotti, nell'individuazione delle metodologie e nelle attività di addestramento e supervisione. 6. Il caso della anonimizzazione. 6.1. Il testo della decisione giurisdizionale come vettore di informazioni. 6.2. Sostenibilità dell'utilizzo dei dati e dell'attività di oscuramento. 7. Profili tecnici dell'impiego dell'IA: sicurezza, spiegabilità e sostenibilità ecologica. 8. Impiego dell'IA in funzione predittiva.

1. Introduzione

Il contenuto di questo documento è inedito, avendo ad oggetto le attività in corso di realizzazione nella introduzione delle tecnologie di intelligenza artificiale nella Giustizia amministrativa, con lo scopo di fornire una iniziale informazione, generale ma sufficientemente esaustiva e diretta, delle evoluzioni avviate dal Segretariato Generale della Giustizia amministrativa, in una fase nella quale l'interesse verso l'impiego di queste tecnologie ha raggiunto una intensità particolarmente accentuata, sia per il livello di diffusione e sviluppo tecnologico sia a seguito delle iniziative di regolazione in ambito unionale e nazionale.

Costanti nelle trattazioni sul tema dell'intelligenza artificiale sono le illustrazioni dei profili giuridici ad esso connessi, con un ruolo centrale della Giustizia amministrativa che per prima ha fornito risposte ad esigenze di tutela legate all'impiego degli algoritmi nei procedimenti amministrativi, all'accessibilità ai codici sorgente, con approfondimenti di analisi estesi anche ad aspetti tecnici concernenti le distinzioni tra varie tipologie di algoritmi, quelli deterministici, costruiti secondo logiche rigidamente causali ("*if-then-else*"), quelli che tali non sono in quanto costruiti secondo logiche probabilistiche e quelli che si basano su modelli di IA, con modalità di apprendimento automatico (c.d. *machine learning),* a loro volta distinguibili a seconda che l'apprendimento sia supervisionato o meno.

Minore evidenza viene data a un altro aspetto, altrettanto noto e non meno rilevante: il ruolo del giudice amministrativo non solo nell'accompagnare, con le proprie pronunce, la modernità ma di artefice, per quanto si andrà ad illustrare, dei processi di innovazione tecnologica nel settore, complesso e delicato, della giustizia. La Giustizia amministrativa, infatti, ha da sempre considerato con lungimiranza l'evoluzione tecnologica quale preziosa opportunità per l'efficientamento dei propri processi interni, con l'obiettivo di assicurare *standard* qualitativamente e quantitativamente sempre più elevati. Così è stato quando per prima tra tutte le giurisdizioni, a gennaio del 2017, ha dato avvio al processo amministrativo telematico, digitalizzato integralmente e non in singole sue fasi, così è attualmente, nell'approccio alle sfide, nuove e ancora per larga parte inesplorate, poste dall'impiego delle tecnologie di IA.

Questa capacità di visione, al tempo stesso libera da pregiudizi e cauta – per imprescindibili esigenze di sicurezza intese nella accezione più ampia, in connessione con i rilevanti valori implicati –, costituisce un *fil rouge* che non si è mai spezzato, grazie al quale la Giustizia amministrativa è nelle condizioni di ponderare con avvedutezza gli sviluppi da percorrere, perché in questo settore come in nessun altro ci sono passaggi evolutivi che non possono essere improvvisati, postulando basi di partenza solide e affidabili.

Con queste premesse, nella prima parte del contributo (par. 1 – 5) verranno illustrati gli impieghi della IA nella Giustizia amministrativa, quelli già consolidati e quelli in corso di sviluppo, l'approccio seguito con declinazione delle sue ricadute concrete, il ruolo essenziale svolto dai giudici amministrativi, in stretta sinergia con la componente ingegneristica, all'interno del Servizio per l'informatica (SPI), nella definizione dell'architettura degli strumenti di IA, nell'individuazione delle metodologie più affidabili, nell'attività di supervisione e addestramento e, in sintesi, nella "istruzione" degli strumenti di IA. Si tratta di qualcosa di molto diverso, complesso e creativo della customizzazione di prodotti tecnologici acquisibili sul mercato, come si cercherà di mettere in evidenza, con un approccio concreto e interdisciplinare, nella seconda parte (par. 6), attraverso l'analisi di uno dei casi d'uso previsti, concernente la anonimizzazione. Lo sviluppo successivo (par. 7) ha ad oggetto profili tecnici di particolare rilievo, riferiti alla sicurezza e alla sostenibilità ecologica della trasformazione digitale in atto. Il documento si conclude con alcuni brevi riferimenti (par. 8) all'impiego dell'IA in funzione predittiva.

2. L'IA nella Giustizia amministrativa: applicazioni consolidate e progetti PNRR

L'utilizzo di tecnologie basate sull'intelligenza artificiale non è una novità assoluta nella Giustizia amministrativa. È il perimetro del campo di applicazione di queste tecnologie che è in rapida evoluzione.

Dal 2020, infatti, nella Giustizia amministrativa sono in uso sistemi avanzati di *cybersecurity* basati sull'intelligenza artificiale che hanno consentito, finora, di preservare il proprio patrimonio informativo da incidenti informatici.

Con l'avvio dei progetti di IA, la Giustizia amministrativa ha dovuto evolvere ulteriormente la propria strategia di sicurezza *cyber* – come si andrà più approfonditamente ad evidenziare nella parte finale del presente contributo –, che è stato particolarmente considerato, "*by design*", nella elaborazione dei progetti di impiego delle tecnologie di IA, con l'introduzione di cautele specifiche secondo le *best practice* migliori suggerite anche da un nuovo settore di ricerca, denominato "Intelligenza artificiale spiegabile" (*Explainable AI o XAI*) che mira a rendere trasparenti i complessi e spesso oscuri algoritmi di apprendimento automatico, fornendo, al contempo, preziosi spunti per delineare una architettura dei prodotti di IA più sicura in relazione agli specifici rischi di sicurezza che la caratterizzano.

La Giustizia amministrativa, infatti, tramite il Servizio per l'Informatica ha attuato progetti altamente innovativi, inseriti nella misura PNRR 1.6.5 ("Digitalizzazione delle grandi PAC – Consiglio di Stato"), che includono la realizzazione di una piattaforma di *business intelligence* e intelligenza artificiale.

Per la realizzazione di questo progetto, la Giustizia amministrativa ha aderito ad un accordo quadro Consip, il cui contratto esecutivo è stato sottoscritto il 22 novembre scorso e ha consentito, nel pieno rispetto del

termine previsto, di disporre, in una prima fase in via sperimentale, di una piattaforma che potrà essere successivamente integrata con il sistema SIGA, implementata e ulteriormente evoluta.

3. Precondizioni dell'avvio dei progetti di impiego delle tecnologie IA e sviluppi prefigurabili

I progetti di impiego delle tecnologie di intelligenza artificiale nella Giustizia amministrativa con le caratteristiche del tutto peculiari che li caratterizzano, affondano le proprie basi su un contesto solido di sviluppo nel quale andranno ad inserirsi.

Sono tre le fondamentali precondizioni che hanno reso possibile l'avvio dei progetti di IA.

La prima: un livello di digitalizzazione già molto elevato, senza il quale non sarebbe stato possibile nemmeno immaginare di approcciare i progetti di intelligenza artificiale.

Dal 1° gennaio 2017 è in vigore il processo amministrativo telematico (PAT) e sono stati digitalizzati anche i fascicoli degli anni precedenti. Questo è un cospicuo vantaggio per l'introduzione dell'IA: i dati di apprendimento sono già tutti digitali e questo *database* costituisce la risorsa, preziosa e sicura, che utilizziamo per addestrare i modelli di IA.

La seconda: scelte infrastrutturali adeguate.

È interessante osservare che nel settembre 2020 la Giustizia amministrativa presentava il proprio progetto di finanziamento in ambito PNRR, includendo, oltre alla piattaforma di *business intelligence* e IA, anche la migrazione dei propri sistemi e applicativi sul *cloud*. L'inclusione della piattaforma tra le progettualità previste è rimasta ferma, mentre la migrazione sul *cloud*, pur essendo stata espunta da queste progettualità, è stata avviata lo stesso anno, con risorse proprie dell'amministrazione. A questo processo è stata impressa una forte accelerazione lo scorso anno: il 9 settembre 2023 il *core* del sistema informativo della Giustizia amministrativa, SIGA, è stato trasferito su *service cloud provider* certificati, in conformità alla disciplina di riferimento.

Da evidenziare è che SIGA non è semplicemente un contenitore organizzato di informazioni non correlate ma una piattaforma integrata in grado di gestire ogni aspetto dell'attività giurisdizionale, consentendo la piena operatività di tutte le categorie di utenti (giudici, segreterie, parti, difensori).

Lo sviluppo epocale di SIGA che è possibile intravedere attraverso l'integrazione con sistemi di IA è quello di un assistente digitale controllato che supporta il giudice, fungendo da "copilota".

Il passaggio su *cloud* è un necessario presupposto per raggiungere ulteriori obiettivi, tra i quali non solo un efficiente sistema di continuità operativa, evitando blocchi dovuti a malfunzionamenti dei *data center*, ma anche la gestione dei progetti di IA e delle loro successive evoluzioni con adeguate risorse computazionali.

Su questo aspetto, si ritiene opportuna una notazione. All'inizio di quest'anno la Corte di Giustizia dell'Unione europea ha pubblicato la sua strategia per l'uso dell'IA. Il documento indica una *best practice* nelle soluzioni IA che devono essere installate e utilizzate *on-premise*, con una valutazione negativa delle tecnologie *cloud* in ambito giudiziario. Al riguardo, il documento non esplicita le argomentazioni tecniche alla base dell'approccio seguito, essendo possibile l'utilizzo di *private cloud* che consentono alla singola

organizzazione il controllo e la gestione interna. Questa indicazione della Corte di Giustizia è probabilmente legata a possibili utilizzi più rischiosi dell'IA, attualmente non previsti dalla Giustizia amministrativa italiana.

In ogni caso quella adottata ha costituito una scelta obbligata: uno dei pilastri della strategia nazionale di trasformazione digitale è il principio *cloud first* e, inoltre, la regolazione intervenuta non consentiva più ai *data center* – come quelli in uso – classificati di "tipo B" in base al censimento condotto dall'AgID nel 2019 di essere ulteriormente potenziati.

La terza precondizione: una forte integrazione delle competenze. Nel Servizio per l'informatica della Giustizia amministrativa opera una componente magistratuale che lavora in stretta sinergia con quella ingegneristica, sia interna sia esterna e questo costituisce un fattore essenziale per gli sviluppi "ordinari" del sistema informativo, e, per quanto si andrà ad esporre, assolutamente decisivo per un'attuazione, accurata e ponderata, dei progetti di IA.

Entro la data del 30 settembre 2024, fissata per la scadenza della progettualità, sono già stati collaudati i *tool* riferiti ai cinque casi d'uso pianificati, ciascuno caratterizzato da un perimetro ben definito di funzionalità, secondo i requisiti stabiliti conformemente agli impegni assunti, con inclusione anche di interfaccia necessarie per la fase di sperimentazione che, avviata successivamente alla suddetta data, diverrà propedeutica alla futura integrazione con il SIGA (e, dunque, all'utilizzazione da parte di tutti gli utenti), in esito alla quale potrà provvedersi a sviluppi evolutivi e integrativi già in larga parte preconizzati.

4. Linee di indirizzo seguite e ricadute pratiche

L'avvio del progetto di realizzazione di una piattaforma di IA ha reso evidente da subito l'importanza di una forte integrazione delle competenze tecnologiche e giuridiche in quanto solo attraverso una stretta sinergia tra competenze multidisciplinari è possibile traguardare gli ambiziosi *target* previsti.

I casi d'uso che sono stati definiti muovono nel solco di quelle che sono state le scelte già operate nel 2020 dalla componente ingegneristica dello SPI.

La linea direttrice seguita con costante impegno dal Segretariato Generale della Giustizia amministrativa e con la supervisione attenta dei vertici dell'Istituto è quella della valorizzazione degli impatti positivi ritraibili dagli sviluppi tecnologici sull'organizzazione del lavoro ma con chiara attribuzione a questa tecnologia di un ruolo strumentale, di supporto al giudice nella fase di studio, aggiornamento e analisi. L'attività di elaborazione resta affidata al giudice in via esclusiva.

Sul piano concettuale risulta più appropriato discorrere di "intelligenza accelerata" invece che di "intelligenza artificiale".

I modelli di *machine learning* – soprattutto di *deep learning* (sistemi, cioè, di apprendimento automatico che simulano, attraverso reti neurali artificiali multistrato, l'azione del cervello umano, come, ad esempio i LLM – *large language models* che utilizzano sistemi di questo tipo a scopi linguistici) sono difficili da governare e possono determinare fenomeni di c.d. di allucinazione o anche fenomeni di *overfitting*, quando il modello si adatta troppo ai dati di addestramento specifici o di *overgeneralization*, quando il modello, all'opposto, generalizza troppo.

Lo scorso anno la Commissione nuove tecnologie della FBE –*Fédération des Barreaux d'Europe*, nella considerazione dei rischi derivanti dall'impiego dell'IA generativa e dai modelli linguistici di grandi dimensioni nel settore legale, ha elaborato delle linee guida, individuando *best practice* ritenute chiave per mantenere alti *standard* etici, salvaguardare la riservatezza, assicurare un utilizzo consapevole e responsabile di queste tecnologie.

Anche di recente le ricerche nel settore (il riferimento è, in specie, a quella svolta dall'Università di Stanford) hanno evidenziato la difficoltà di questi modelli di IA ad eseguire il tipo di ragionamento giuridico, individuando vari fattori di incidenza negativa, inclusa la mancanza di uniformità e la lunghezza delle frasi presenti nei documenti legali.

L'approccio non poteva, dunque, che essere cauto, consapevole dei rischi e al tempo stesso libero da pregiudizi ma caratterizzato da spirito critico e da una supervisione costante.

Al centro restano le persone perché:

- le funzioni giurisdizionali sono per Costituzione affidate al giudice persona fisica, naturale e precostituito per legge, terzo e imparziale, soggetto solo alle leggi;
- i valori etici e la tutela dei diritti non possono essere delegati alla tecnologia;
- la capacità di *leadeship* non può essere automatizzata.

Né va trascurata la consistenza delle risorse necessarie per lo sviluppo di questi progetti.

Nella fase iniziale di introduzione di queste tecnologie si è pensato, quindi, a una base di partenza che non vincola rispetto alle scelte future e alle strategie di impiego della IA preconizzabili, necessariamente soggette nel tempo ad aggiornamenti e riassetti. Per contro, se questi progetti non fossero stati avviati la Giustizia amministrativa avrebbe perso l'opportunità anche di scegliere efficacemente il futuro che corre in questo settore talmente velocemente da rendere difficile recuperare il tempo trascorso.

In linea con l'approccio seguito nell'impiego delle tecnologie di IA nella Giustizia amministrativa, sono state effettuate scelte specifiche.

- **esclusione dell'utilizzo dell'intelligenza artificiale per la generazione di testi di qualsiasi tipologia**. Si tratta di sviluppi certamente possibili su di un piano astratto e tecnico ma che sono connotati da elevata complessità e delicatezza nel settore in riferimento, richiedendo anche capacità computazionali e risorse molto consistenti;
- **i casi d'uso previsti non rientrano tra gli usi definiti ad alto rischio secondo l'IA Act,** approvato in via definitiva dal Parlamento europeo il 13 marzo 2024 e, successivamente, dal Consiglio il 21 maggio 2024, con pubblicazione nella Gazzetta Ufficiale dell'Unione europea avvenuta in data 12 luglio 2024 (regolamento UE 2024/1689 del 13 luglio 2024). I casi d'uso, infatti, non riguardano l'attività decisionale del caso concreto che resta affidata esclusivamente al giudice. Le applicazioni previste mirano a fornire esclusivamente supporto nello svolgimento di attività preparatorie, organizzative o di approfondimento. Al riguardo, si segnala anche l'approvazione, il 23 aprile 2024, da parte del Consiglio dei Ministri di un disegno di legge, n. 1146 AS, avente ad oggetto "Norme per lo sviluppo e l'adozione di tecnologie di intelligenza artificiale", che legittima l'impiego delle tecnologie di IA nel settore

giudiziario solo per l'organizzazione e la semplificazione del lavoro e per la ricerca giurisprudenziale e dottrinale, essendo stato opportunamente espunto, rispetto al testo preliminare informalmente circolato su *internet*, il riferimento alla possibilità di usare l'IA per predisporre anche solo le bozze dei provvedimenti giurisdizionali, specificandosi, altresì, che è sempre riservata al magistrato la decisione sulla interpretazione della legge, sulla valutazione dei fatti e delle prove e sulla adozione di ogni provvedimento. Anche rispetto a questa iniziativa legislativa, emerge la piena conformità delle progettualità in corso di realizzazione da parte della Giustizia amministrativa;

- **i dati di addestramento dei modelli adottati sono costituiti dai dati presenti nel *database* interno**, non provengono da fonti esterne o "aperte", con rafforzamento, quindi, della sicurezza;
- **per ogni caso d'uso è stata progettata una *pipeline* specifica**. Questo aspetto merita una particolare considerazione in quanto si tratta di una architettura esclusiva, essendo le *pipeline* disegnate congiuntamente dai giudici e dal *team* ingegneristico, composto da personale interno e fornitori esterni selezionati mediante procedura pubblica. Le *pipeline* consentono di attuare un approccio controllato e supervisionato che evita il fenomeno della c.d. *black box*, facilitando un processo nel quale ogni azione viene "catturata" per essere "spiegata". In questa architettura un ruolo fondamentale è affidato alle attività umane che consentono di "plasmare" le tecnologie impiegate in stretto ancoraggio con le finalità perseguite e nella prospettiva della maggiore sicurezza e affidabilità di un risultato che deve, comunque, essere sempre verificato in un processo di miglioramento costante;
- **non viene mai utilizzata un'unica tecnologia ma una pluralità di metodi e modelli** in combinazione tra loro (regole algoritmico-procedurali; *large language models*; *named entity recognition* e varie tecniche di elaborazione del linguaggio naturale). Ciò è molto utile sia sul piano delle *performance* sia per ragioni di sicurezza, poiché l'impiego di più tecnologie rende più difficile la riuscita di eventuali attacchi;
- **non è supervisionata solo l'attività di addestramento iniziale ma anche il *fine tuning* successivo** per un periodo congruo e non breve. Ciò significa che il *feedback* fornito dagli utenti in fase di utilizzo dei prodotti non verrà recepito automaticamente dai modelli ma avverrà solo dopo l'analisi e la valutazione di un'apposita unità, composta dagli ingegneri e dai magistrati che stanno lavorando allo sviluppo dei progetti;
- **attenzione per l'ambiente in termini di risparmio energetico** ma con diretta incidenza anche sui costi. Ciò avviene attraverso *algoritmi di quantizzazione* che consentono di trasformare dati ad alta dimensionalità in uno spazio compresso senza compromettere significativamente le prestazioni. Questo processo permette:
    - minor consumo di energia;
    - occupazione di minori risorse e quindi minori costi;
    - minore latenza e maggiore velocità.

Non si sta trascurando nemmeno questo aspetto perché una Corte *smart* non è quella che persegue l'efficienza a tutti costi ma quella che bilancia accuratamente i diversi valori in gioco.

A quanto esposto deve aggiungersi una ulteriore riflessione. La realizzazione dei progetti di IA non deve indurre a trascurare i processi evolutivi che attengono ai sistemi e agli applicativi già in uso nella Giustizia amministrativa e, cioè, quelli che costituiscono l'attività consolidata, la quale è in continuo sviluppo, richiedendo costanti interventi, con lo scopo del perseguimento di *standard* qualitativi e quantitativi sempre più elevati nell'erogazione dei servizi offerti, fondamentali per i singoli e per la collettività.

Se ciò non può essere revocato in discussione, devono però anche essere sottolineati due aspetti, non di poco conto. In prospettiva, infatti, e cioè quando si addiverrà all'integrazione dei *tool* di IA con SIGA, tutte le tecnologie disponibili convergeranno, sinergicamente, nella direzione dei risultati attesi; inoltre, è possibile concepire sviluppi degli strumenti già in uso (si pensi, ad esempio, alla complessa attività di reingegnerizzazione dei portali esistenti) sulla base di soluzioni che rendano più efficace l'impiego delle tecnologie più innovative. Il riferimento è, in specie, all'utilità che sarebbe ritraibile da una uniformità dei *format* degli atti processuali, secondo l'esperienza già maturata in altre Corti (Corte di Giustizia dell'Unione europea, Corte europea dei diritti dell'uomo, altre Corti di Stati dell'UE), quale fattore che renderebbe possibile, a prescindere da ulteriori e fondamentali obiettivi (effettività dei principi di chiarezza e sinteticità), una più efficace introduzione delle tecnologie IA.

5. Casi d'uso e ruolo del giudice nella definizione dell'architettura dei prodotti, nell'individuazione delle metodologie e nelle attività di addestramento e supervisione

I casi d'uso delle tecnologie di IA realizzati possono essere ripartiti in due macro categorie:

- quelli che sono funzionali ad agevolare varie attività del giudice, nei diversi ruoli, e che, per come strutturati, non prevedono un'attività "creativa", essendo diretti a efficientare e velocizzare attività di ricerca e di rilevazione e visualizzazione di contenuti;
- il caso d'uso concernente la anonimizzazione dei provvedimenti giurisdizionali, che presenta aspetti più delicati, in prospettiva condizionando anche, in parte, uno degli altri casi d'uso (quello concernente la ricerca dei precedenti giurisprudenziali).

Nella prima categoria sopra indicata rientrano:

a) l'**identificazione di ricorsi correlati o simili pendenti nelle singole Sezioni** e che devono essere fissati per la decisione.

   Individuare ricorsi simili su cui decidere, ad esempio quelli che riguardano le stesse questioni giuridiche, consente di raggiungere diversi obiettivi:

   - ottimizzazione dello studio e dell'analisi;
   - valutazione ai fini della discussione nella stessa udienza o in udienze "tematiche";
   - evitare decisioni contrastanti nelle singole sezioni dei Tribunali;
   - ottenere una migliore distribuzione dei carichi di lavoro;
   - garantire decisioni più rapide.

Si tratta, quindi, di un caso d'uso utile per il personale dell'Ufficio del processo, delle Segreterie e per i Presidenti, al fine di migliorare il processo di individuazione delle cause da fissare per la trattazione in udienza;

b) la **ricerca dei precedenti giurisprudenziali** con uno strumento basato non solo, come ora, su parole chiave bensì sulla rilevazione di connessioni semantiche, in questo modo garantendo un maggior grado di pertinenza dei risultati della ricerca;

c) la **rilevazione e possibilità di visualizzazione immediata delle norme o delle pronunce della giurisprudenza** (costituenti due casi d'uso autonomi sebbene assimilabili sia per finalità perseguite che per tecnologie impiegate) indicate, esplicitamente o implicitamente, in un atto difensivo, evitando al giudice di dover interrompere l'analisi dell'atto per svolgere la ricerca su banche dati esterne, in tal modo assicurando risparmio di tempo ed evitando anche che la sua concentrazione venga distolta.

Per quanto riguarda la **anonimizzazione**, invece, il *target* è quello di uno strumento che, ferma restando l'attività di verifica, modifica e validazione del personale delle Segreterie, fornisca una proposta di anonimizzazione che sia il più possibile conforme alla normativa di riferimento e al tempo stesso eviti gli eccessi, di sovente registrati nello svolgimento di questa attività da parte delle Segreterie, che compromettono la stessa intellegibilità dei provvedimenti e ne rendono poi difficile se non impossibile la ricerca.

L'attività di customizzazione, il c.d. *tailor made*, è una attività ordinaria quando si tratta di prodotti informatici che devono essere adattati alle specifiche esigenze di una organizzazione, ai suoi processi, alla disciplina normativa di riferimento.

Nel caso dei progetti di IA non si tratta di questo per il modo in cui la Giustizia amministrativa li sta attuando.

I magistrati che lavorano a questi progetti non si limitano a orientare meri adattamenti di prodotti esistenti ma contribuiscono all'individuazione del metodo, fornendo un apporto essenziale per la creazione della stessa architettura di ciascuno di questi prodotti che saranno una esclusiva della Giustizia amministrativa perché pensati e disegnati dalla Giustizia amministrativa.

Lo sforzo maggiore nell'attuazione di questi progetti è quello della integrazione delle competenze, dovendo il giudice proporre ma anche semplificare concetti giuridici - che non sono sempre nettamente definiti – declinandoli in processi logico - scientifici, per rendere possibile la loro traduzione in regole informatiche.

Traslare in ambito matematico una scienza dello spirito, come tradizionalmente viene considerato il diritto, è attività estremamente complessa perché ci si deve sforzare di superare, attraverso regole che devono essere via via affinate e controllate nei risultati, fisiologiche ambiguità per giungere a un *output* che sia il migliore possibile.

Anche se nessuno dei casi d'uso che la Giustizia amministrativa ha attuato prevede la generazione di testi, sin da subito è emersa con particolare evidenza la delicatezza di questi strumenti.

Nonostante i tempi brevi a disposizione per la realizzazione della progettualità PNRR, è stato da subito scartato un approccio non controllato all'impiego dell'IA: si tratta di un approccio che consente all'IA di sviluppare processi di comprensione e decisione senza linee guida o regole rigide.

L'intervento del giudice nell'attuazione di questi progetti – come si andrà ad approfondire nel seguito della trattazione – risulta cruciale:

- nella definizione delle regole di base iniziali e cioè nel processo tecnicamente definito di *prompt engineering,* nel quale sostanzialmente l'IA viene istruita sul contesto, le descrizioni, i vincoli da rispettare;
- nel *design* per ogni caso d'uso di una architettura degli strumenti da impiegare in modo tale da rendere il processo completo, coerente ed "*explainable*" in ogni sua fase;
- nella verifica dei risultati, che vengono analizzati anche al fine del perfezionamento delle regole;
- nella supervisione e nel *fine tuning* successivo;
- nella mappatura e raccolta delle regole che si vanno a definire perché oltre all'intelligenza artificiale dovrà essere "istruita" anche quella umana, evitando disallineamenti, nel senso che gli utenti dovranno essere informati per poter correttamente utilizzare questi strumenti.

Si tratta di un confronto costante tra diverse competenze coinvolte in questi progetti, che includono anche esperti di linguistica, che si sta rivelando molto proficuo grazie all'elevatissima qualificazione *in primis* del personale ingegneristico del Servizio per l'informatica della Giustizia amministrativa.

Dall'esperienza che si sta maturando, toccando con mano, entrando nei meccanismi di funzionamento di questi processi, risulta allo stato estremamente difficile immaginare, anche nel medio termine e in disparte i vincoli normativi, l'introduzione di un *tool* di IA che sia in grado di elaborare bozze di provvedimenti con livelli minimi di garanzia che nel settore giudiziario sono irrinunciabili.

Non si deve, infatti, commettere l'errore di "sovrapporre" i modelli linguistici, per quanto larga possa essere la loro dimensione, alla decisione, sul presupposto che il pensiero, anche quello giudico, si esprime con il linguaggio; è evidente, infatti, che simili semplificazioni sono ardite e scadono nella banalizzazione per la constatazione, ovvia, che non è sufficiente il linguaggio per articolare un ragionamento giuridico pertinente e coerente. Questi approdi sono ancora in fase di elaborazione a livello tecnologico e scientifico e la loro introduzione non è seriamente ipotizzabile in tempi brevi per ragioni legate anche alle capacità computazionali di cui la Giustizia amministrativa può concretamente disporre, al netto dell'approccio al quale si intenda aderire su di un piano anche etico.

Più fattori spingono verso un approccio aperto e al tempo stesso cauto, evitando eccessi di entusiasmo nonostante rispetto al passato due fattori determinanti hanno impresso la spinta alla quale stiamo assistendo: la disponibilità di grandi quantità di dati e la rilevante potenza di calcolo. Pur con questi punti di forza resta il fatto che l'IA non deve produrre diritto, restando il ragionamento e l'elaborazione giuridica appannaggio esclusivo del giudice. Se questo è vero e se, dunque, l'IA non sostituirà il giudice, è innegabile che i giudici che utilizzeranno, nei limiti e con le cautele dovute, queste tecnologie acquisiranno un valore aggiunto e già questo rende il senso del perché questi fenomeni non devono essere subiti ma governati e conosciuti nella prospettiva di un servizio Giustizia sempre più evoluto ed efficiente.

6. Il caso della anonimizzazione

L'obiettivo di questa parte del presente documento è descrivere come, nel concreto, si proceda alla generazione di un sistema di intelligenza artificiale in ambito giuridico e come questo venga poi progressivamente addestrato. In particolare, ci si soffermerà sull'oscuramento dei dati sensibili all'interno di una sentenza e in generale negli atti giudiziari, a partire dalle premesse teoriche sino a giungere alle principali implicazioni operative.

Tra i possibili approcci, si è preferito rifarsi ad argomenti e regole tratti da ambiti scientifici ed essenzialmente dalla teoria dell'informazione, osservando il fenomeno dell'oscuramento come un'attività fisica diretta a rimuovere una parte dei dati disposti e organizzati nel testo, secondo un preciso ordine logico e secondo l'imprescindibile funzione esplicativa della decisione.

Non ci si propone dunque di esaminare le regole giuridiche che presiedono all'oscuramento, quanto semmai di osservare le implicazioni operative del loro uso, del loro apprendimento e della loro evoluzione da parte di un sistema di intelligenza artificiale, destinato ad assolvere due compiti fondamentali: impedire la diffusione di dati sensibili, così da proteggere la riservatezza degli interessati; assicurare la comprensibilità del testo della decisione, garantendo, senza alterarla se non attraverso la rimozione delle informazioni sensibili, l'intellegibilità della motivazione posta alla base della decisione giurisdizionale.

Si tratta quindi di bilanciare due esigenze di segno opposto, risolvendone il conflitto.

Di seguito si affronteranno, pur con taglio essenziale, alcuni meccanismi che amplificano tale conflitto, le possibili strategie dirette ad attenuarlo e limitarne gli effetti, nonché l'addestramento e lo sviluppo dell'intelligenza artificiale impiegata per governarlo.

6.1. Il testo della decisione giurisdizionale come vettore di informazioni

Quando redige una sentenza, il giudice dialoga con le parti attraverso una sorta di narrazione, descrivendo fatti e spiegando le conseguenze che desume da quei fatti, sino a proporre loro uno scritto originale. Collega le informazioni alle regole, secondo il proprio convincimento e la propria scienza, deducendo una o più affermazioni costitutive che interferiscono con la realtà giuridica.

Dal punto di vista che qui interessa, l'obiettivo del giudice è organizzare i dati in modo chiaro e coerente per rendere comprensibili le premesse e il contenuto della decisione. I dati, in quest'ottica, sono informazioni organizzate che servono a fornire la chiave di lettura di tale contenuto, ovvero informazioni (sensibili o no) che sono introdotte nel testo per assolvere una funzione esplicativa.

Muovendo da una prospettiva opposta e speculare, ci si può interrogare su cosa significhi ottenere una sentenza in cui sono stati oscurati i dati sensibili (per la loro definizione in ambito giudiziario il riferimento va rivolto alla disciplina dell'Unione Europea e, in particolare, al GDPR).

In termini strettamente materiali, si può osservare che oscurare significa predisporre una copia dello scritto originale. Solo che in questa copia sono eliminati i dati coperti da riservatezza (cd. dati sensibili).

Viene quindi utilizzato un procedimento inverso rispetto a quello seguito dal giudice, che ne rovescia il lavoro e fa sì che aumenti l'incertezza dei dati rispetto alla versione originale, ossia rendendo il testo meno comprensibile.

Questo aumento dell'incertezza è il risultato dell'eliminazione di informazioni specifiche, operazione che, tuttavia, intacca anche il significato complessivo, risultante dalla connessione delle informazioni localizzate all'interno del testo.

Con l'attività di oscuramento, il dilatarsi dell'incertezza non ha quindi a che fare soltanto col valore informativo di isolatati dati specifici, ma con la relazione che, in base alla loro posizione e alla loro connessione nel testo, essi hanno con gli altri. È, in effetti, all'interno di questa relazione che si genera un valore informativo ulteriore, contribuendo a delineare tutte le premesse giuridico-fattuali della decisione.

Il dove (con riguardo sia all'analisi topografica sia alla struttura logica del testo) sono inserite le informazioni può così accrescere in modo rilevante la conoscenza. Pertanto, la posizione delle informazioni rimosse contribuisce ad amplificare il *deficit* di conoscenza generato dal loro oscuramento, moltiplicandolo rispetto al valore in sé di ciascuna informazione eliminata.

Quando l'oscuramento compromette le informazioni interconnesse, la decisione perde la gran parte del proprio significato perché ciò priva l'argomentazione giuridica dalla necessaria premessa concreta.

Tenendo ferma questa conclusione, si può osservare che se, dunque, la stesura della decisione impone di strutturare le informazioni, attribuendo loro una funzione esplicativa, rimuoverle, attraverso l'oscuramento, significa, invece, intaccarne la struttura.

Ma fino a che punto può essere compromessa la chiarezza della decisione? Fino a quale soglia può essere ritenuta inevitabile e quindi irrisolvibile la perdita di informazioni che, come visto, scaturisce dall'oscuramento dei dati sensibili e dalla loro destrutturazione?

Al riguardo, deve essere sottolineato che un'IA genericamente addestrata è tutto sommato in grado di oscurare uno scritto, ma ciò non è però sufficiente per la risoluzione del problema e per la conservazione della funzione esplicativa della decisione giudiziaria. Il sistema può essere forzato a distinguere le informazioni in base al loro valore semantico: nomi propri; sentenze penali e titoli delittuosi; quelle molteplici entità semantiche che definiscono malattie, infermità, orientamenti religiosi e sessuali, condizioni economiche, ecc. Ma quello che l'IA non può fare o che non riesce a fare ancora in modo efficiente, basandosi essa esclusivamente sull'utilizzo di modelli linguistici e sull'approccio statistico che le appartiene (che tende a divenire sempre meno affidabile quanto più la trama delle informazioni si faccia complessa e dilatata), è intercettare le relazioni all'interno del testo e quelle premesse concrete che si antepongono all'argomentazione.

In effetti, come si è accennato, distinguere il processo decisionale dallo scritto che lo rappresenta consente di guardare a quest'ultimo come uno strumento esplicativo della decisione. Perciò, con l'oscuramento non possono essere eliminate indiscriminatamente tutte le informazioni strutturate ma solo alcune, pena la compromissione della struttura e dunque dell'efficacia chiarificatrice del testo. Ne consegue che la funzione della procedura è, in ultima analisi, creare un certo grado di limitata incertezza, bilanciando la chiarezza e la comprensibilità del testo organizzato con la protezione della riservatezza e dei dati sensibili. Questa funzione

si innesta, conformandola, nell'attività addestrativa, situata nella fase preparatoria, attività che costituisce, in effetti, uno dei fattori che contribuisce a conferire maggiore precisione ed equilibrio all'impiego dell'IA nel caso d'uso in esame, e richiede necessariamente l'intervento dei magistrati, il cui compito comprende la collaborazione con la componente ingegneristica nella predisposizione di un insieme di istruzioni - coerente con le regole giuridiche - dirette ad infondere adeguata incertezza alle sole informazioni sensibili senza, contemporaneamente, alterare il senso complessivo del testo della decisione.

Il raggiungimento di un equilibrio tra i due aspetti (protezione della riservatezza e comprensibilità del testo) presuppone infatti un'attività interpretativa che può essere svolta solo da giudici, i quali, in tale fase, contribuiscono a definire il valore intrinseco da assegnare a ciascuna informazione, considerandone anche il rapporto con il contesto, distinguendo la natura e le caratteristiche delle controversie.

L'intelligenza artificiale si modella così assecondando la variabilità del contesto, per poi verificare nel corso dell'utilizzo concreto, grazie ai *feedback* ricevuti dagli utenti (*feedback* opportunamente verificati e somministrati alla stessa IA), la correttezza dei risultati ottenuti ancora sotto il controllo dei giudici, i quali, in tale fase esecutiva, sorvegliano lo sviluppo del sistema ossia il modo in cui esso risponde ai riscontri offertigli dall'utenza, vigilandone il percorso evolutivo.

6.2. Sostenibilità dell'utilizzo dei dati e dell'attività di oscuramento

Le considerazioni che precedono inducono a svolgere alcune riflessioni riguardo alla sostenibilità dell'uso dei dati e dell'attività di oscuramento, ponendo in relazione il lavoro di organizzazione delle informazioni, proprio della redazione della sentenza, e il contrapposto lavoro – non più eseguito dal giudice – necessario per rimuoverne una parte.

È possibile minimizzare (e fino a che punto) il dispendio energetico associato all'oscuramento dei dati sensibili?

La risposta dipende essenzialmente (oltreché dall'efficienza del processo di oscuramento) da quanto, nella fase di redazione della sentenza, sia stato sorvegliato e quanto più possibile ridotto il numero di informazioni sensibili utilizzate, in ragione della loro effettiva utilità nel definire le ragioni della decisione.

Se, in effetti, durante la redazione della decisione il giudice utilizzasse il minor numero possibile di informazioni sensibili, nella fase successiva di oscuramento dei dati sensibili, si dovrebbero conseguentemente rimuovere o modificare meno informazioni e sarebbe quindi minore (ovvero quanto minore possibile) il dispendio energetico necessario per oscurare i dati sensibili.

Questo approccio, del resto, sembra profilare anche l'emersione di un vantaggio aggiuntivo in termini di sicurezza dei dati, poiché riducendo il numero di informazioni sensibili presenti nella versione originale dello scritto, viene anche diminuito il rischio associato alla loro potenziale divulgazione o compromissione.

In definitiva, si può osservare e concludere che, anche in un'ottica di sostenibilità, minimizzare il numero di informazioni sensibili nei provvedimenti giurisdizionali potrebbe essere considerata una strategia efficace per ridurre lo spazio di archiviazione, il lavoro e le risorse necessari all'intelligenza artificiale per oscurare i

dati sensibili successivamente, riducendo nel contempo anche il dispendio energetico e l'uso delle risorse, nonché contribuendo a migliorare (non va dimenticato) gli *standard* di sicurezza dei dati.

7. Profili tecnici dell'impiego dell'IA: sicurezza, spiegabilità e impatto ambientale

L'impiego della IA nella Giustizia amministrativa non è del tutto inedito, in quanto, come in precedenza evidenziato, dal 2020 vengono utilizzate queste tecnologie nella sicurezza *cyber*, costantemente rafforzate e implementate.

Nell'ambito della sicurezza informatica, l'IA consente di esaminare rapidamente grandi volumi di dati, di individuare attività anomale con accuratezza e di rappresentare le stesse in forma agevolmente comprensibile all'utente finale, in tal modo accrescendo la reattività negli incidenti informatici, con riduzione dei tempi di contrasto alle intrusioni.

In particolare, la peculiarità di analizzare grandi quantità di dati consente, attraverso appositi strumenti di monitoraggio *realtime*, di ricavare le informazioni che servono sui principali *malware*.

Nello specifico, l'IA può fornire un sostanziale aiuto suggerendo le azioni da compiere:

- per prevenire un attacco da uno specifico *malware*;
- immediatamente dopo aver individuato una risorsa compromessa dal *malware* in questione e più in generale, per individuare quale sarà il prossimo *step* seguito dal *malware*. In quest'ultimo caso l'IA consente di intervenire *ex post* su quanto già accaduto per attuare una *remediation* specifica ed *ex ante* su quanto potrà avvenire consentendo l'implementazione di una corretta politica di prevenzione.

L'IA, più in generale, consente:

- un'agevole individuazione delle minacce mediante il monitoraggio delle attività all'interno della rete;
- la prevenzione di incidenti di sicurezza, evitando le intrusioni non autorizzate e proteggendo i dati sensibili contenuti nei dispositivi;
- l'attuazione di misure correttive e adeguative a seguito di attacchi.

Non deve però essere trascurato il rapporto bidirezionale esistente tra IA e sicurezza informatica.

Da un lato, l'IA può aiutare a rilevare e prevenire le minacce informatiche, a verificare l'identità degli utenti, a monitorare e controllare le reti. Dall'altro lato, l'IA può essere usata per creare nuove forme di attacco informatico, più sofisticate e difficili da contrastare, in un contesto nel quale la minaccia è in crescente aumento.

Dal report 2023 pubblicato da Clusit, confrontando i dati del 2018 con quelli del 2022, emerge un incremento del 60% degli attacchi rilevati (da 1.554 a 2.489). Secondo M-Trends 2023, il 63% degli incidenti informatici sono registrati da entità esterne all'organizzazione.

Sono state individuate alcune tecniche di attacco specificamente rivolte ai sistemi informatici basati su IA, denominate *adversarial machine learning*, riconducibili essenzialmente alle seguenti tre categorie:

- *data poisoning attack:* mira a influenzare i dati utilizzati nell'addestramento o nel riaddestramento di un modello di IA. Una porzione di dati contaminati viene immessa nel *training-set* del modello in modo che esso impari nel modo sbagliato. La contaminazione può consistere nell'associazione di una

label sbagliata oppure nella polarizzazione verso una categoria specifica di *input*. In tal modo dati dannosi (contaminazione) vengono erroneamente classificati come legittimi, portando così al rifiuto di dati legittimi dopo l'addestramento. Ad esempio la compromissione del filtro *antispam* di un fornitore di posta elettronica rientra nella prima categoria. Nel caso in questione riferito a *gmail*, gli attaccanti hanno segnalato in maniera massiccia *email* legittime come *spam*, in tal modo, una volta compromesso il filtro *antispam*, gli aggressori sono stati in grado di inviare varie *e-mail* dannose, che includevano *malware* e altre minacce, aggirando efficacemente i filtri di sicurezza senza essere intercettati. Altro esempio è quello dell'attacco contro il sistema *mobileye* di Tesla che ha indotto l'auto a guidare 80 km/h oltre il limite di velocità semplicemente aggiungendo una striscia di nastro nera di circa 5 cm a un segnale di limite di velocità;

- *evasion attack*: *a*ttraverso un *malware* si cerca di compromettere le previsioni di un modello di AI: con l'introduzione di impercettibili modifiche agli *input* originali si induce il modello a fare errori di classificazione. Un interessante esempio di applicazione di questo tipo di tecniche è legato all'abbigliamento utilizzato (ad es. particolari maglie che non coprono il volto) per eludere le tecniche di riconoscimento facciale. Approcci simili possono essere applicati in altri contesti, come nell'analisi dei *malware*;
- *extraction attack: i*n questo tipo di attacco l'obiettivo è quello di recuperare il *training set* su cui il modello di IA è stato addestrato. La dimensione del problema è aumentata se del *training set* fanno parte dati sensibili o riservati violando, in tal caso, la normativa sulla *privacy*.

Le tecniche "*adversarial*" sono quindi in primo luogo un pericolo per i sistemi difensivi, in quanto permettono di aggirare allarmi e strumenti di controllo di accesso che sono stati appresi in base ai dati disponibili. Possono però divenire anche uno strumento difensivo, nel senso che possono servire a progettare sistemi di *anomaly detection* e di autenticazione che siano più robusti rispetto a questo tipo di attacchi.

L'attenzione costantemente elevata riservata nella Giustizia amministrativa alla sicurezza informatica ha ricevuto ulteriore impulso proprio con l'avvio dei più recenti progetti di impiego delle tecnologie di IA. La progettazione dell'architettura di IA con l'approccio *security by design* consente, già sul nascere, un'attenta prevenzione delle minacce suindicate al fine di definire una architettura che sia il più possibile sicura e affidabile.

In linea con le analisi svolte a livello nazionale e internazionale, la direzione verso la quale si sta operando muove dalla consapevolezza che l'idea di *cybersecurity* non è più "sufficiente", essendo necessario un approccio più ampio e trasversale, quello della *cyber*-resilienza: la capacità di continuare a fornire i risultati attesi nonostante il verificarsi degli attacchi informatici.

Vengono, inoltre, costantemente considerate le specifiche strategie di difesa dalle tecniche di *adversarial machine learning* individuate a livello internazionale, tra le quali:

- il rinforzo del *training*: inserendo nell'addestramento esempi avversari si espone il modello a *input* perturbati costringendolo a imparare a generalizzare meglio non facendo affidamento su caratteristiche specifiche;

- l'*ensemble* di modelli: costruire un *ensemble*, cioè combinare le previsioni di diversi modelli, può rendere più difficile per un attacco influenzare tutti i modelli contemporaneamente. Gli *ensemble* aumentano la diversità delle risposte, migliorando la resistenza complessiva agli attacchi;
- la verifica dei dati: implementare controlli di verifica per rilevare esempi avversari nel *dataset* di addestramento può aiutare a ridurre la probabilità che tali esempi influenzino il modello. La rimozione di esempi avversari dal *dataset* o l'etichettatura accurata di tali esempi può preservare l'integrità dell'addestramento;
- l'*input transformation*: applicare trasformazioni leggere agli input durante la fase di inferenza può aiutare a ridurre l'impatto degli attacchi avversari. Queste trasformazioni, come il rumore casuale aggiunto all'*input*, possono rendere più difficile per gli attacchi generare perturbazioni efficaci;
- l'analisi di sensibilità: condurre analisi di sensibilità per identificare le regioni dello spazio di *input* in cui il modello è più vulnerabile agli attacchi può guidare la progettazione di difese mirate e consentire un miglioramento della robustezza;
- aggiornamenti continui: l'*adversarial machine learning*, infatti, è un settore in continua evoluzione.

L'implementazione di una combinazione delle tecniche e strategie illustrate, infatti, può aiutare a mitigare i rischi associati all'*adversarial machine learning* e a garantire che i modelli di *machine learning* siano più resistenti agli attacchi informatici.

Si evidenzia, inoltre, che tra le difficoltà di approccio ai sistemi di IA, anche nel campo della *cyber* sicurezza, vi è la difficoltà nel comprendere il funzionamento di tali sistemi che appaiono delle vere e proprie *black-box*.

Recentemente un nuovo settore di ricerca, denominato Intelligenza Artificiale Spiegabile (*Explainable* AI o XAI), cerca di fare chiarezza sull'argomento con lo scopo di rendere trasparenti i complicati e spesso oscuri algoritmi di apprendimento automatico. Nonostante i molteplici approcci recentemente emersi per spiegare le decisioni dei classificatori *black-box*, essi non appaiono sempre di agevole e intuitivo utilizzo per gli utenti finali. Per gli esperti di IA è difficile capire come *l'output* venga generato in modelli altamente complessi di tipo *data driven*, tramite algoritmi che coinvolgono milioni di parametri tra loro interagenti. La natura intrinsecamente complessa di questo processo rende difficile per gli esperti comprendere appieno come questi modelli producano i loro *output* finali.

Gli sviluppi delle analisi svolte nell'ambito della XAI costituiscono oggetto di costante attenzione per vagliare l'introduzione delle metodologie più idonee ad assicurare la spiegabilità dei sistemi di IA dei quali è prevista l'introduzione nella Giustizia amministrativa.

Rilevano, al riguardo, i recenti approfondimenti che, al fine di rendere chiaro un modello di *deep neural network* (DNN) evidenziano la centralità di due aspetti: l'interpretazione o interpretabilità e la spiegabilità. Questa duplice attenzione all'interpretabilità e alla spiegabilità contribuisce a una migliore comprensione delle complessità coinvolte nel funzionamento dei modelli DNN (usati nelle applicazioni di IA della Giustizia amministrativa). L'interpretabilità consente agli sviluppatori di esplorare il processo decisionale del modello, migliorando così la comprensione delle modalità con cui il modello produce i suoi risultati. A

differenza di una semplice previsione, la tecnica dell'interpretazione consente di ottenere informazioni aggiuntive o spiegazioni cruciali per comprendere il funzionamento sottostante di un sistema di intelligenza artificiale fornendo, a chi è in possesso di adeguate competenze, un aiuto ulteriore per una migliore comprensione del modello *black-box*. Al contrario, la spiegabilità consente all'utente finale di acquisire fiducia nell'accuratezza e nell'obiettività dell'IA, fornendo dettagli sul processo decisionale della DNN.

L'obiettivo generale è quello di porre gli utenti "umani" in condizione di comprendere, di fidarsi maggiormente e di supervisionare con competenza le comunità artificialmente intelligenti le quali, inoltre, sono in continua evoluzione.

Su tali premesse l'architettura di IA sviluppata per la Giustizia amministrativa tiene in forte considerazione i concetti di supervisione e controllo dei modelli di apprendimento automatico. Benché la piattaforma in fase di realizzazione, attraverso i suoi servizi di base, sia in grado di erogare servizi di intelligenza artificiale avanzata, con la possibilità di usufruire di modelli preaddestrati di alta specializzazione, in grado di compiere nativamente funzionalità molto complesse, per garantire un controllo e una configurabilità della soluzione che sia a tutti i livelli ed al massimo dettaglio possibile, l'approccio seguito è stato impostato su elevati livelli di controllo.

Ciascuno dei casi d'uso previsti viene implementato realizzando delle catene di elaborazione (*pipeline*) molto granulari, organizzate quindi in diversi passi (*step*) con compiti ben determinati dal punto di vista logico e funzionale. Attraverso diverse fasi di verifica, gli *output* dei vari *step* sono esaminati, al fine di accettarne la regolarità dell'elaborazione. In aggiunta, il sistema di controllo e gestione generale monitora il flusso dall'acquisizione del dato fino alla restituzione del risultato al fine di garantire la gestione del processo, il controllo dei dati e la sicurezza del sistema.

I principali vantaggi di questo approccio possono essere riassunti nei seguenti punti:

- flessibilità, in quanto le *pipeline* sono ideate come configurabili;
- controllo, ogni operazione e configurazione della *pipeline* è tracciata e supervisionata;
- spiegabilità, i modelli di IA forniscono le spiegazioni delle loro risposte;
- accuratezza, vengono tracciate e valutate le metriche delle varie operazioni;
- modularità, per garantire la gestione/evoluzione, riuso e il controllo;
- ottimizzazione, ad esempio: vengono selezionati i modelli più performanti e efficienti in termini di risparmio energetico (LLM quantizzati).

L'architettura di IA, sviluppata all'interno del Servizio per l'informatica della GA, prevede specifiche fonti di dati come *input* per poter implementare i differenti *use case*. I *database* individuati per il raggiungimento dei *target*/*milestone* PNRR sono il *database* operazionale ed il *database* documentale della Giustizia amministrativa.

Nello strato di elaborazione dell'architettura sono gestiti differenti modelli di IA, scelti e utilizzati rispetto alle loro specializzazioni. Tipicamente modelli linguistici più "leggeri" (NER) svolgono le funzioni di riconoscimento delle entità all'interno di un testo mentre i *Large Language Model* (LLM) che, tecnicamente

possono essere configurati come DNN, svolgono funzioni più complesse come la classificazione e il riconoscimento semantico del testo.

Il processo principale che governa il sistema è il processo di "*deploy*", può essere riassunto nei seguenti passi:

- *preprocessing* dei dati: in questa fase, i dati raccolti dalle diverse fonti vengono puliti, trasformati e preprocessati per renderli adatti all'addestramento del modello. Ciò può includere operazioni come la rimozione di valori mancanti, la normalizzazione dei dati e la codifica delle variabili categoriche;
- selezione delle caratteristiche: si identificano le caratteristiche rilevanti presenti nei dati che saranno utilizzate per addestrare il modello. Questo passo è cruciale per migliorare le prestazioni del modello e ridurre il tempo di addestramento;
- addestramento del modello: in questa fase, si selezionano e si addestrano i modelli di intelligenza artificiale utilizzando algoritmi di apprendimento supervisionato, non supervisionato o *reinforcement learning*, a seconda del problema specifico da affrontare. Durante l'addestramento, il modello impara dai dati e ottimizza i suoi parametri per minimizzare l'errore nelle previsioni;
- validazione e ottimizzazione del modello: una volta addestrato il modello, si valuta la sua accuratezza e le sue prestazioni utilizzando metodi di validazione come la *cross-validation* o l'insieme di *test*. In base ai risultati, si può ottimizzare ulteriormente il modello mediante la selezione di parametri, la regolarizzazione o *l'ensembling*;
- caricamento del modello: si carica il modello addestrato precedentemente salvato dal componente di sviluppo e addestramento;
- *preprocessing* dei dati in tempo reale: Si applicano le stesse trasformazioni e il *preprocessing* dei dati utilizzati durante l'addestramento ai nuovi dati in *input* per garantire la consistenza e la corretta interpretazione delle informazioni;
- inferenza: si utilizza il modello addestrato per effettuare previsioni o classificazioni sui nuovi dati in *input*;
- *postprocessing*: si elaborano le previsioni o le classificazioni ottenute dal modello, ad esempio convertendo i risultati in un formato comprensibile per gli utenti finali o integrandoli con altre informazioni.
- esposizione dei risultati: si forniscono i risultati dell'elaborazione del modello agli utenti finali o ad altri componenti del sistema, ad esempio attraverso API, interfacce utente o *dashboard*.

Una particolare attenzione va rivolta alle modalità con cui i dati sono conservati e gestiti all'interno del sistema (persistenza dei dati). Le tipologie di persistenza usate sono:

- *vector db*: dedicato alla memorizzazione di dati in forma vettoriale, che è essenziale per l'implementazione di diversi servizi IA. L'archiviazione dei dati in forma vettoriale è importante per eseguire le operazioni vettoriali in maniera efficiente sfruttando le librerie di *machine learning* disponibili. I vettori possono essere memorizzati in strutture dati ottimizzate, come *array* o tensori, che facilitano la manipolazione e l'elaborazione pre-elaborata dei dati da parte degli algoritmi di apprendimento automatico;

- *database* relazionale o RDBMS: è dedicato a fornire tutte le strutture di tipo tecnico, necessarie all'esecuzione degli algoritmi, come ad esempio le tabelle anagrafiche ed al monitoraggio della piattaforma come le tabelle di *logging* e di esecuzione;
- *object storage*: sistema di archiviazione scalabile e affidabile utile per l'archiviazione e il recupero di oggetti in modo trasparente rispetto all'infrastruttura del sistema, al *backup* per l'archiviazione a lungo termine e per la gestione dei cicli di vita dei dati e alla condivisione di contenuti in modo sicuro, fornendo URL firmati o impostando le politiche di accesso.

Non trascurabili, infine sono la componente di controllo/gestione e quella relativa ai risultati (*output*). La prima raggruppa e fornisce diverse funzionalità come la gestione, il controllo e la sicurezza. Alcune di queste funzionalità non sono concentrate in un unico applicativo ma sono distribuite funzionalmente su più *target* (ad esempio per la sicurezza). Più nel dettaglio le principali componenti possono essere individuate nelle seguenti:

- schedulazione: adibita all'esecuzione organizzata dei processi della piattaforma, in particolare per i processi di tipo '*batch'*;
- metriche: si occupa della raccolta elaborazione ed organizzazione delle *performance* della piattaforma;
- stato: designata a monitorare e verificare lo stato di funzionamento della piattaforma e delle sue componenti;
- sicurezza: racchiude tutte le funzionalità e applicazioni che si occupano di garantire in tutti gli aspetti la sicurezza della piattaforma;
- *log*: si occupa di raccogliere e analizzare i diversi log prodotti dalla piattaforma, ed ha come scopo primario il *trouble shooting* dei processi;
- *Dev/Ops ML/Ops* è il sottosistema che si occupa del ciclo di vita degli algoritmi e dei modelli della piattaforma.

I presidi elevati di sicurezza attuati in fase di realizzazione dei progetti di introduzione della IA nella GA, anche associati alle misure generali di protezione dei sistemi in uso, per quanto accurati ed evoluti, trovano, però, necessario complemento nel comportamento degli utenti, risultando imprescindibili cautele in una duplice prospettiva.

Da un lato, infatti, sul piano tecnico e generale, tutti i sistemi di protezione, a partire dell'autenticazione multifattoriale (MFA), lungi dal dover essere percepiti quale appesantimento delle attività, rientrano ormai tra le regole di base alle quali deve essere assicurata conformità.

Dall'altro, con specifico riferimento alle tecnologie di IA, per quanto numerosi siano i controlli dei sistemi e anche del *fine tuning* successivo alla loro realizzazione, resta fermo il controllo che rientra tra le responsabilità proprie di ogni utente, in relazione alle attività da loro svolte, in conformità, del resto, al principio della riserva di umanità.

Sebbene, infatti, i casi d'uso in precedenza illustrati non rientrino tra quelli considerati ad alto rischio in base all'IA Act e non pongano problematiche di portata analoga a quelle legate ad un uso più "spinto" dell'IA, come ad esempio, quello di generazione di testi, restano sempre ferme le responsabilità che da sempre connotano lo svolgimento delle attività, specie in un settore delicato quale è quello della giustizia.

Un ulteriore profilo merita di essere evidenziato e attiene alla rilevanza dell'impatto ambientale che l'impiego dell'IA determina e all'attenzione dedicata anche a questo profilo nella realizzazione dei progetti della Giustizia amministrativa.

Le transizioni ecologica e digitale, fortemente interconnesse, sono spesso considerate "gemelle", per la rilevanza strategia che entrambe assumono a livello unionale e nazionale.

Riflettere sull'impatto energetico delle tecnologie *disruptive* come il *cloud* e l'intelligenza artificiale risulta fondamentale e le scelte attuate nella Giustizia amministrativa valorizzano questi aspetti.

I *cloud provider* utilizzati, selezionati in conformità alla normativa di riferimento, effettuano consistenti investimenti in tecnologie di alimentazione e raffreddamento all'avanguardia, progettate per consumare meno energia, rendendo meno impattante l'elevato numero di *server* che gestiscono. Inoltre, l'utilizzo di servizi in *cloud* si basa su risorse condivise, come reti, macchine di calcolo e strutture fisiche, pensate per essere sfruttate a pieno, riducendo gli sprechi, potendo contare, tra l'altro, sulla ottimizzazione per la scalabilità dinamica della frequenza di voltaggio, risparmiando risorse. In sintesi, i *provider* di servizi in *cloud* possono realizzare grandi economie di scala, massimizzando l'ottimizzazione e l'utilizzo di *hardware* e processi efficienti. Né va trascurata l'adozione di fonti di energia rinnovabile per alimentare le infrastrutture digitali, specie associata all'impiego di reti intelligenti decentralizzate in grado di gestire l'aumento delle energie rinnovabili e di distribuirle secondo le esigenze.

Con riferimento alle tecnologie di IA, notoriamente energivore, nelle progettualità della GA si è fatto ricorso, come già accennato, a un processo specifico, di quantizzazione, che determina una riduzione della precisione dei dati senza però comportare rilevanti ricadute sulla significatività degli stessi, con il vantaggio di una contrazione dei consumi energetici.

La quantizzazione consiste nel convertire i pesi e le attivazioni delle reti neurali da numeri in virgola mobile a numeri interi con larghezza di *bit* inferiore. Sebbene possa, in linea teorica, influenzare la precisione dei modelli di *machine learning* (specie nei modelli che usano il *deep learning*), è spesso necessaria per ridurre la quantità di memoria e la potenza di calcolo e, di conseguenza, ridurre l'impatto energetico.

Il ricorso a tale processo è complesso in quanto implica un accurato bilanciamento tra efficienza e accuratezza per garantire prestazioni ottimali nelle applicazioni di intelligenza artificiale, con contestuali risparmi sia in termini di impatto ambientale sia di costi.

8. Impiego dell'IA in funzione predittiva

Si ritiene, infine, per completezza, di fare riferimento a un ulteriore profilo.

Gli sviluppi tecnologici alimentano costantemente il dibattito in ordine ad un possibile ruolo della giustizia predittiva.

Invero, è sulla stessa accezione di "giustizia predittiva" che si dovrebbe riflettere, perché se l'espressione, intesa in senso letterale, potrebbe essere intesa come capacità di un sistema di IA di elaborare una decisione, s*trictu sensu* indica, invece, la possibilità di risalire rapidamente, attraverso queste tecnologie, alla disciplina

pertinente per trattare un caso e contestualizzarlo secondo specifiche caratteristiche anticipando la probabilità delle decisioni che potrebbero essere prese.

In questa accezione la giustizia predittiva ambisce a rendere l'applicazione del diritto più prevedibile, favorendo l'accesso all'informazione, affinando l'elaborazione scientifica, offrendo alle parti analisi sul "rischio giudiziario" allo scopo di valutarlo senza offrire la "soluzione" ma diversi possibili scenari e la loro probabilità di verificarsi.

Nessuno dei casi d'uso in precedenza illustrati include un'applicazione con lo scopo di predire l'esito di una determinata decisione.

In senso lato, il caso d'uso relativo alla ricerca dei precedenti della giurisprudenza, potrebbe avere una funzione orientante rispetto al possibile esito di una decisione. Ma non si tratta di un impiego direttamente mirato a tale scopo.

L'impiego della IA in chiave predittiva è stato, invece, previsto per altre finalità e con un differente strumento.

È stato realizzato, infatti, il progetto di realizzazione del *datawarehouse*, che cambierà radicalmente le capacità di analisi statistica della GA. Lo scopo è introdurre processi di *business intelligence* evoluti, costantemente aggiornati ed efficaci.

Si tratta di un *repository* centrale di dati e metadati che vengono processati attraverso dei *data mart* che consentono diverse modalità di aggregazione, in base a criteri predeterminati. Questo sistema può permettere, tra l'altro, di conoscere in tempo reale:

- numero delle pendenze nel complesso e in ogni singolo Ufficio giudiziario;
- tempi di definizione delle controversie in primo grado e in appello;
- la dimensione non solo quantitativa e ma anche qualitativa del contenzioso.

Lo strumento è utile per la *governance* e per orientare in maniera informata le decisioni sul piano organizzativo.

Dopo la prima fase di creazione del *datawarehouse*, attualmente utilizzato in via sperimentale, si è passati alla seconda fase, attualmente in corso, funzionale a rendere possibili previsioni maggiormente attendibili sulle tendenze in atto e sulle proiezioni future.

Anche questo progetto è stato realizzato con fondi PNRR ed è stato, nella sua prima fase, concluso prima della scadenza prevista, con rendicontazione già positivamente eseguita a livello europeo, con apprezzamento espresso per contenuti e per le concrete modalità di realizzazione.

Brunella Bruno
Responsabile del Servizio per l'informatica
Nicola Bardino
Vicario del Servizio per l'Informatica e Responsabile del trattamento dei dati personali
Domenico Franco Sivilli
Direttore generale per le risorse informatica e la statistica della Giustizia amministrativa