

INTELLIGENZA ARTIFICIALE E SANITÀ: I PROGETTI DI ARIA SpA

La società in house della Regione Lombardia ha introdotto un percorso di sviluppo delle attività di AI partendo dallo studio delle cronicità tra i cittadini. Obiettivo: analizzare, in ottica predittiva, l'insorgere di diverse patologie. I progetti però non si fermano qui.

Vincenzo Virgilio



Giuseppe Preziosi, responsabile del Data Warehouse e Business Intelligence Competence Center di ARIA

Individuare con anticipo il numero complessivo di persone che, in un determinato periodo di tempo, potranno essere affette da una determinata patologia cronica. È possibile farlo, utilizzando strumenti di intelligenza artificiale. La dimostrazione arriva dalla nuova Azienda Regionale per l'Innovazione e gli Acquisti (ARIA SpA), nata il 1° luglio dalla fusione di ARCA, la Centrale Acquisti regionale, e di Lombardia Informatica, la digital company della Regione, alle quali si aggiungerà a breve anche ILSPA (Infrastrutture Lombarde). Nel suo ruolo di supporto strategico all'analisi, all'elaborazione e all'utilizzo del patrimonio informativo – funzione, questa, ribadita come centrale anche nel recente emendamento n. 160 alla Legge Regionale di assestamento di bilancio – ARIA (già Lombardia Informatica) negli ultimi anni ha posto un accento importante alla valorizzazione delle informazioni sanitarie istituendo un proprio Data Warehouse e Business Intelligence Competence Center.

La struttura, istituita dal 2017 con l'obiettivo di convergere nel processo di trasformazione digitale, ha sviluppato a oggi diverse iniziative di AI "con l'obiettivo di programmazione, di controllo della spesa, di valutazione dei percorsi diagnostico terapeutici dei cittadini della regione e di prevenzione del rischio", ci ha spiegato **Giuseppe Preziosi**, il responsabile del centro di competenza che abbiamo intervistato per capire come l'intelligenza artificiale possa essere un supporto alla governance in ambito sanitario e avere un quadro dell'esperienza di Lombardia Informatica finora realizzata in questo campo.

Quali funzioni ha all'interno di ARIA il centro di competenza di cui è responsabile?

L'obiettivo del Competence Center è sistematizzare una 'value proposition' fondata sull'impiego e l'elaborazione del patrimonio informativo di ARIA attraverso un modello di offerta incentrato sui dati, sulle informazioni e soprattutto sulle conoscenze da queste generabili. Il nostro focus al momento è l'healthcare e in questo ambito ARIA ha costruito e visto crescere nel tempo il Data Warehouse del Welfare, in cui sono raccolti ed elaborati più di dieci anni di dati sanitari derivati principalmente dai flussi amministrativi – relativi, per esempio, a ricoveri ospedalieri, consumo di farmaci, erogazione di prestazioni ambulatoriali – di oltre 10 milioni di cittadini lombardi, con un modello organizzativo che nel tempo ha visto uno 'shift' dal mero possesso dei dati al loro utilizzo. Siamo nell'era del dato. Per la prima volta nella storia moderna è la tecnologia a inseguire i dati e non i dati a cercare una tecnologia che li supporti e li generi. Per farlo, dal 2017 abbiamo sperimentato anche l'applicazione di strumenti di intelligenza artificiale con un modus operandi che ha seguito quanto già suggerito nel libro bianco dell'AI stilato dall'Agenda per

l'Italia Digitale (AgID), che prevede un primo approccio sperimentale della tecnologia per poi estenderla alle caratteristiche peculiari dell'organizzazione, dei dati e della mission aziendale. Si tratta ancora di una fase di studio e sperimentazione di gran fermento, prodroma a traguardare il nostro obiettivo di rinascere da crisalide dopo un processo di data driven transformation.

Quali progetti avete sviluppato in ambito sanitario e con quali obiettivi?

La sfida del sistema sanitario, come un Giano bifronte, è quella di bilanciare la domanda, ossia il bisogno, con l'offerta, ossia con le capacità produttive finite. Partendo da questo presupposto, abbiamo voluto occuparci di un aspetto dirompente nell'attuale quadro epidemiologico che è la crescita della cronicità, con patologie che interessano circa il 30% della popolazione lombarda con un dispendio del 70-75% delle risorse, in perfetta logica paretiana.

Secondo la Delibera di Giunta Regionale nr. X/6164 sulla Presa in Carico degli assistiti cronici lombardi: "[...] Dopo una prima fase [...] il modello di presa in carico è, infatti, destinato ad evolvere secondo un approccio bottom-up, che valorizzi i modelli predittivi [...]. In sintesi i cittadini lombardi possono essere classificati, in base alla loro domanda di salute, su cinque livelli di stratificazione. [...] La stratificazione della domanda consente anche di ripensare in modo proattivo il sistema regionale della prevenzione, al fine di anticipare la domanda di salute con idonee azioni personalizzate [...]". L'ambito prescelto di applicazione delle metodiche innovative di AI è quello della clusterizzazione della popolazione assistita, da sempre oggetto di analisi, valutazione ed evoluzione in Lombardia – basti citare le esperienze della Banca Dati Assistito e la stratificazione della Delibera della Domanda – ma basate finora su elementi ex post (esenzioni, consumi di farmaci, prestazioni ambulatoriali e di ricovero). In questo contesto, si anticipa l'innovazione introdotta da tecniche di machine learning in ottica previsionale e anticipatoria.

Come avete operato e quale tipo di cronicità avete messo al centro di questo progetto?

L'iniziativa è partita con il progetto 'Predictive BDA', ossia l'evoluzione in ottica predittiva della Banca Dati Assistito della Regione, che finora si è occupata di stratificare la popolazione lombarda, in funzione dei propri consumi, su 62 patologie croniche per traccianti, livello di complessità, territorio, genere, fascia di età e assistenza. Il progetto ha visto la direzione scientifica dell'Osservatorio Epidemiologico della Regione Lombardia guidato dal dottor Luca Merlini della Direzione Generale Welfare.

Nel corso del 'proof of concept' abbiamo voluto circoscrivere il fenomeno oggetto di studio a un perimetro specifico, quello relativo alle cardiopatie, intese come pa-

tologie croniche a cui afferiscono l'ipertensione arteriosa, la cardiopatia ischemica e lo scompenso cardiaco e che rappresentano un ramo particolarmente significativo tra le cronicità sia in termini di diffusione nella popolazione a livello di incidenza percentuale sui residenti e di numero totale di pazienti interessati, che in termini di assorbimento delle risorse economiche e della spesa indotta, tipiche variabili di analisi del campione nell'ambito della totalità delle patologie croniche. I dati quantitativi sono tutti disponibili sul sito open data dell'Osservatorio Epidemiologico (<https://www.dati.lombardia.it/osservatorio-epidemiologico>) nella sezione 'Il modello lombardo per la presa in carico'. Sono state prese in considerazione le informazioni degli ultimi tre anni di consumi relativi all'analisi degli assistiti, all'acquisto di farmaci, alle prestazioni ambulatoriali, alle esenzioni dei cittadini e ai ricoveri ospedalieri, per una porzione di popolazione pari a circa due milioni di persone, ossia il 20% di assistiti lombardi. In sintesi, il modello

Per la prima volta nella storia moderna è la tecnologia a inseguire i dati e non i dati a cercare una tecnologia che li supporti e li generi.

predittivo è stato fondato, attraverso tecniche di machine learning, sulla capacità di prevedere il numero di persone che, sane a una certa data, avrebbero potuto sviluppare una cardiopatia come patologia cronica prevalente nell'anno successivo.

Dal punto di vista privacy e sicurezza delle informazioni, i dati presi in considerazione nel progetto sono stati pseudonimizzati e

sono state adottate ulteriori misure di generalizzazione per impedire la riconduzione all'individuo, nonché di sicurezza per consentire gli accessi ai soli data scientist accreditati, minimizzando qualsiasi rischio di elaborazione o esfiltrazione non autorizzata.

Quali sono stati i risultati?

Il modello che abbiamo studiato nel 2017 ha previsto un numero di cronici con un errore pari allo 0,04% rispetto al numero complessivo di quelli che poi sono stati i cronici effettivamente rilevati: laddove il sistema ne ha stimati 25.065, quelli osservati sono stati 25.055. Lo stesso modello è stato applicato successivamente anche a ritroso, negli anni precedenti, senza tararlo nuovamente, e anche in questo caso abbiamo ottenuto un'accuratezza della previsione dal 96,22% al 97,96%. Livelli senza dubbio apprezzabili anche in funzione dell'obiettivo di supporto alla governance e alla programmazione sanitaria grazie alla rilevazione della spesa indotta e che ci hanno convinto a proseguire in questa tipologia di progettualità ampliandone il raggio d'azione.

Avete sviluppato altre esperienze di AI in ambito sanitario?

Stiamo lavorando su almeno altre due iniziative sempre in stretta condivisione con gli obiettivi di miglioramento della capacità di governance regionale e le finalità di programmazione, gestione, controllo e valutazione dell'assistenza sanitaria. La prima riguarda lo studio, attraverso tecniche

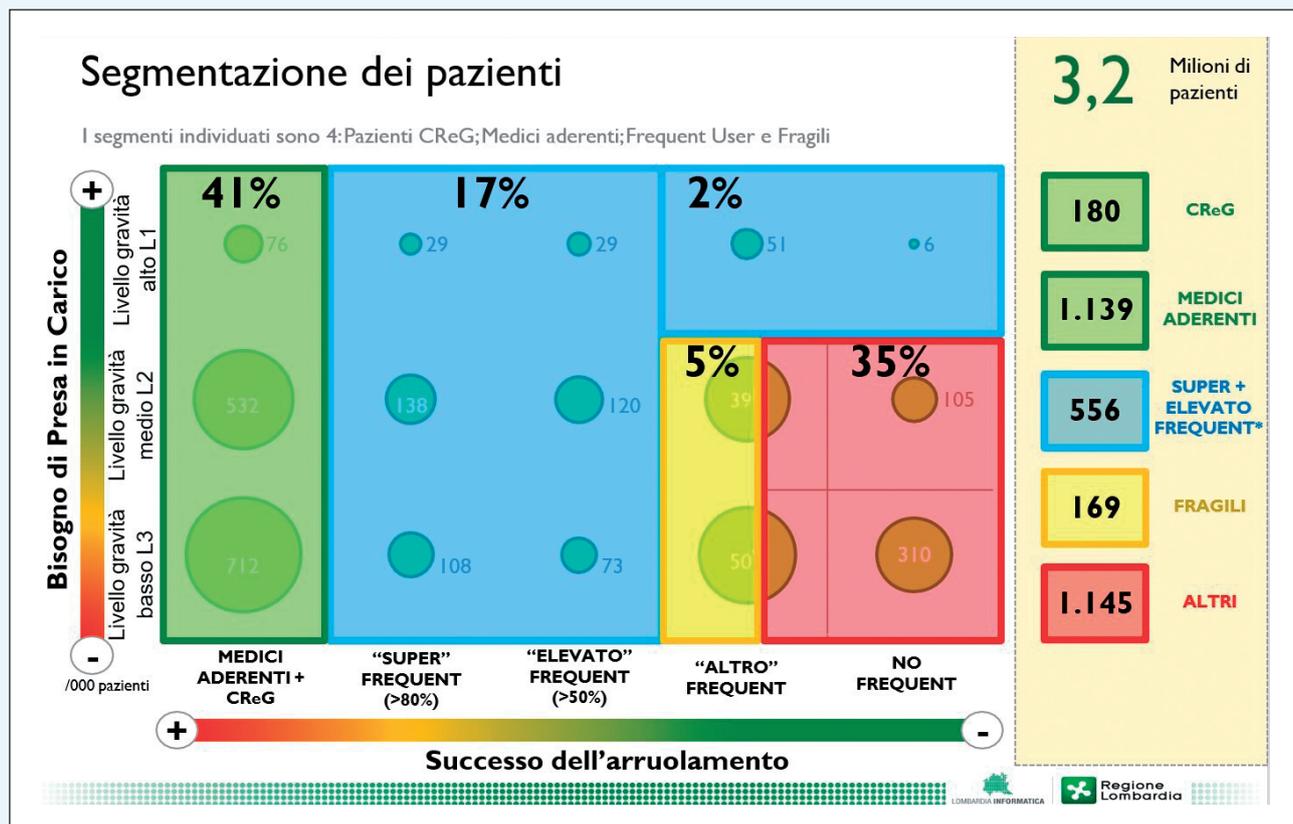
di data & process mining, delle caratteristiche di una determinata patologia: la demenza cognitiva. Il progetto mira a individuare la cosiddetta 'next-best-clinical-action', cioè di analizzare e prevedere, in funzione dei consumi e degli eventi sanitari della popolazione cronica affetta da tale patologia, quale può essere il percorso sanitario, il 'patient journey', compiuto da un assistito cronico, anno per anno, nel quinquennio successivo all'insorgenza della patologia. Per farlo, anche in questo caso ci siamo basati sui traccianti definiti dall'Osservatorio Epidemiologico della Direzione Generale Welfare e abbiamo analizzato un'area territoriale ristretta osservando un fenomeno complesso, basti pensare al fatto che in un quinquennio si possono registrare fino a più di 400 eventi sanitari diversi per ciascun cronico. È un approccio all'AI vicino a quello empirista di John Locke, contrapposto alla scuola razionalista di Gottfried Leibniz. Il primo, empirista, applica metodi statistici per costruire, da grandi quantità di dati, modelli predittivi, algoritmi di classificazione, funzioni di regressione, con un apprendimento 'evidence-based' che va dall'evidenza empirica alla concettualizzazione. Il secondo, razionalista, adotta invece procedimenti ipotetico-deduttivi, usando gli strumenti

della logica, dell'ontologia, della semantica, per costruire sistemi di ragionamento con un apprendimento che va dai concetti (della scienza o del linguaggio ordinario) ai dati. Un'altra iniziativa che abbiamo avviato è quella che abbiamo definito 'machine learning geografico'.

Cosa intende per 'machine learning geografico' e in cosa consiste questa iniziativa?

In questo progetto prendiamo in esame i pazienti affetti da patologie respiratorie, in particolare l'insufficienza respiratoria, la broncopneumopatia cronica ostruttiva (BPCO) e l'asma. In particolare, applicando tecniche di machine learning riusciamo a determinare quella che abbiamo definito UMTA o Unità Minima Territoriale di Aggregazione. Una sorta di area geografica dove sono aggregati assistiti caratterizzati da proprietà sanitarie omogenee, ancora relative ai traccianti sopra citati definiti dall'Osservatorio Epidemiologico Regionale. Una volta determinato questo poligono di Voronoi (decomposizione di uno spazio metrico per la rappresentazione di un determinato fenomeno, ndr) di informazioni sanitarie aggregate – e quindi non riconducibili all'individuo – è possibile procedere

Il modello che abbiamo studiato nel 2017 ha previsto un numero di cronici con un errore pari allo 0,04% rispetto al numero complessivo di quelli che poi sono stati i cronici effettivamente rilevati: laddove il sistema ne ha stimati 25.065, quelli osservati sono stati 25.055.



La segmentazione dei pazienti cronici per il progetto 'Predictive BDA' di ARIA

alla correlazione di ulteriori informazioni del territorio considerato, come potrebbero essere le caratteristiche socio demografiche e quelle ambientali, per poi costruire e applicare un modello previsivo a due anni che va a stimare l'insorgere di patologie di tipo respiratorio. In questo modo si consente alle istituzioni di poter intervenire per tempo e quindi con più efficacia, con le azioni e le politiche di sostenibilità e prevenzione più idonee.

È previsto che i progetti di AI avviati da ARIA possano evolvere e toccare anche ambiti diversi da quelli attuali?

Il contesto di applicazione dell'intelligenza artificiale è decisamente esteso e rispecchia senza dubbio anche una delle caratteristiche peculiari dei progetti in tale ambito, in cui dal nostro punto di vista è necessario un approccio, non solo di 'cognitive intelligence' – l'intelligenza artificiale è identificata dall'Università di Stanford come "una scienza e un insieme di tecniche computazionali che vengono ispirate dal modo in cui gli esseri umani utilizzano il proprio sistema nervoso e il proprio corpo per sentire, imparare, ragionare e agire" – ma anche e soprattutto di 'collective intelligence'.

Nel percorso di trasformazione digitale che ci vede coinvolti nell'adozione di metodiche e tecnologie prossime all'AI, vedremo sempre di più accanto alle figure più tecniche, ingegneristiche, informatiche e di programmazione, profili con competenze di business, legal e magari anche umanistiche, legate alla filosofia e alla psicologia, necessarie per ottenere risultati più efficaci. È questo l'approccio che stiamo cercando di adottare anche in Lombardia Informatica e nel nostro centro di competenza.

Far sedere allo stesso tavolo il clinico, il programmatore, l'informatico, il business analyst, l'esperto di privacy e di security, per avviare processi che non sono solo di digitization, ma di digitalization.

Digitization e digitalization, qual è la differenza e come si traducono questi concetti in medicina?

Digitization in ambito sanitario significa analizzare un documento cartaceo (per esempio un referto, una diaria infermieristica a bordo letto, etc.) e trasporlo in formato elettronico con un processo che porta a disporre di dati, questa volta, strutturati. Digitalization, invece, vuol dire dare valore aggiunto a quei dati strutturati così raccolti, per esempio attraverso l'applicazione dell'AI, con tecniche di speech recognition, natural language processing e machine learning a supporto delle diagnosi o dell'interpretazione delle informazioni.

Il sentiment diffuso sul mercato dell'AI è assolutamente

positivo; non è un caso che tutti i principali Paesi industrializzati stiano definendo una strategia per i programmi di applicazione dell'AI.

È oltretutto già possibile immaginare di programmare sistemi intelligenti che leggono i dati risultanti da esami diagnostici e li interpretano sulla base delle statistiche condotte su grandi moli di dati 'simili', con un processo di miglioramento dell'apprendimento continuo. Esistono addirittura strumenti predittivi per valutare potenziali rischi di evoluzione delle stesse patologie. La medicina di precisione, la ricerca sui trattamenti personalizzati quasi tailor-made ovvero sartoriali, le analisi sui possibili effetti di trattamenti sanitari e farmacologici su singoli pazienti sono tutti campi in fermento e che promettono grandi risultati, sulla carta, grazie alla capacità dell'intelligenza artificiale di immagazzinare ed elaborare molte informazioni, ma che al contempo richiedono sia un accurato controllo da parte della comunità medica, sia una stretta osservanza della privacy e delle norme etiche e deontologiche, mantenendo una prospettiva human-in-the-loop dove è sempre l'essere umano al centro.

Quale potrebbe essere invece un approccio sbagliato all'intelligenza artificiale?

La trappola è la 'sindrome del Silver Bullet', ossia pensare che l'AI sia la soluzione unica che quasi magicamente risolve tutti i problemi.

Basti pensare che nel 2013, l'MD Anderson Cancer Center di Houston, Texas, ha lanciato un progetto 'moon shot': diagnosticare e raccomandare piani di trattamento per alcune forme di cancro utilizzando un sistema cognitivo leader di mercato. Nel

L'AI non è una sfera magica, può produrre errori interpretativi che potrebbero portare a danni significativi, soprattutto in ambito sanitario. È quindi fondamentale mantenere competenze tecniche e tecnologiche, investendo anche in corsi di formazione e puntando a disporre delle figure professionali più idonee.



La struttura del sistema sanitario lombardo

2017, però il progetto è stato 'congelato' dopo che i costi avevano superato i 62 milioni di dollari e il sistema doveva ancora essere utilizzato sul campo. Allo stesso tempo, il gruppo IT del centro oncologico, attraverso iniziative molto meno ambiziose, come una formula per suggerire hotel e ristoranti alle famiglie dei pazienti, ha ottenuto risultati molto più promettenti con un evidente aumento della soddisfazione dei pazienti. Cionondimeno l'MD Anderson continua tuttora a impegnarsi nell'utilizzo di tecnologie cognitive per migliorare il trattamento del cancro.

L'AI inoltre non è una sfera magica, può produrre errori interpretativi che potrebbero portare a danni significativi, soprattutto in ambito sanitario. È quindi fondamentale mantenere competenze tecniche e tecnologiche, investendo anche in corsi di formazione e puntando a disporre delle figure professionali più idonee per le quali, soprattutto per una società in-house, è fondamentale la partnership tra pubblico e privato. Sia dal punto di vista formativo che dell'aggiornamento professionale ci guardiamo attorno e cerchiamo di apprendere dai migliori in questo campo.

Il connubio pubblico-privato, dal punto di vista tecnologico, come si traduce? Avete dei fornitori di riferimento?

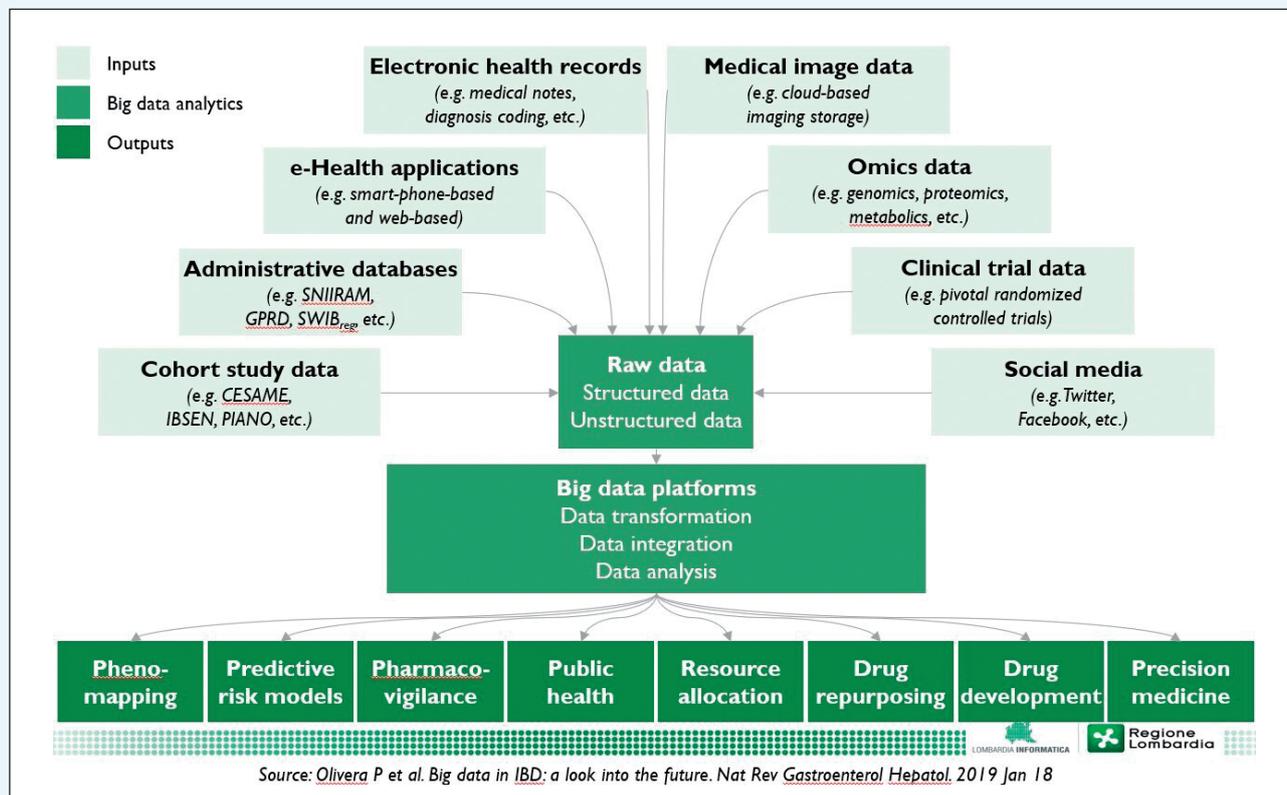
Ci avvaliamo di diversi partner che selezioniamo tramite gare a evidenza pubblica cui partecipano i principali player del mercato. Nello specifico del nostro competence

center è stata indetta una gara su servizi e progetti che interessano anche l'ambito dei sistemi direzionali e del data warehouse, aggiudicata a un raggruppamento temporaneo di imprese che ha dimostrato di possedere le expertise necessarie su questo perimetro.

Cerchiamo comunque di non collaborare con un unico player di riferimento anche perché secondo noi lo strumento, soprattutto in ambito AI, è neutrale: non è infatti la tecnologia a caratterizzare un'iniziativa di intelligenza artificiale, quanto le competenze e i dati sui quali la tecnologia viene applicata. In questo una società in-house deputata alla governance dei dati e delle informazioni svolge un ruolo primario e, quale responsabile del trattamento di quei dati, contribuisce a rendersi garante del loro utilizzo in osservanza delle finalità previste. Come sempre le regole ci sono e sono ben scritte; la nuova regolazione GDPR non è solo a tutela del diritto fondamentale di ciascuno ma anche dell'interesse sociale e pubblico a un corretto trattamento dei dati di tutti: un ampliamento di prospettiva fondamentale che in campo sanitario di ricerca per finalità di tutela della salute trova uno dei banchi di prova più significativi.

È questo il maggior elemento differenziante i progetti di AI rispetto a quelli IT classici?

Dal mio punto di vista esistono due elementi particolarmente differenzianti. Il primo, è la necessità di compliance



Modello 'as a Service' per integrare dati strutturati e non

alla normativa, in particolare al GDPR, perché è chiaro che anche da un punto di vista etico, soprattutto in ambito sanitario, è fondamentale fare avanzare i progressi tecnologici proteggendo al tempo stesso il diritto alla riservatezza e alla dignità del paziente.

Il secondo aspetto è proprio il concetto di 'collective intelligence'. Ci vuole un'importante capacità di leadership per utilizzare la conoscenza 'deep' della realtà e decidere con una visione più estesa attraverso percorsi fra loro interconnessi.

Nelle competenze AI conterà la natura umanistica e comunicativa perché tra le professioni emergenti spiccheranno ruoli chiamati a estendere la 'cultura del dato e dell'analisi' in tutta l'organizzazione aziendale e a disegnare sistemi e interfacce per semplificare e rendere naturale l'interazione uomo-macchina, la cosiddetta touch-point humanization.

Quali nuove figure professionali saranno centrali nella conduzione di progetti di intelligenza artificiale?

I data curator, coloro che operano prevalentemente sulla preparazione dei dati per il loro utilizzo da parte di algoritmi di AI; i machine learning data scientist, che avranno le competenze per 'addestrare' correttamente il sistema e fornire analisi e interpretazione degli outcome; i data evangelist, che avranno invece il compito di 'educare' l'intera organizzazione al corretto utilizzo dei dati e dei sistemi di AI.

Ancora adesso una parte significativa dei progetti di intelligenza artificiale consiste proprio nel porre la 'domanda giusta' a partire dai dati e procedere quindi con le fasi di data preparation e data ingestion, in modo da comprendere dai dati quali sono le variabili più significative e adatte a spiegare il fenomeno di interesse, applicando poi tecniche di machine learning al fine di estrarre gli insight più significativi. Insight che spesso sono inattesi. Sposando questo approccio di 'collective intelligence', il nostro obiettivo è anche quello di creare una sorta di ecosistema di 'Data Information Hub' (DIH) operando in sinergia con i diversi stakeholder.

Per realizzarlo, tre anni fa abbiamo costituito una piattaforma, definita DaaS, Data-as-a-service, che rappresenta anche un modello organizzativo condiviso con la Regione Lombardia e oggetto di alcune delibere e di un'interazione con l'Autorità Garante della Privacy, grazie alla quale vengono messi a disposizione, in una sandbox segregata e ad accesso controllato, i dati sanitari presenti nel Data Warehouse e gli strumenti di data analysis per la loro elaborazione alle Università e agli Enti di ricerca, garantendo la compliance alla base giuridica e il rispetto delle misure di sicurezza ritenute più idonee.

Non è la tecnologia a caratterizzare un'iniziativa di intelligenza artificiale, quanto le competenze e i dati sui quali la tecnologia viene applicata.

AI e sanità: come saranno indirizzati i vostri sforzi nei prossimi mesi e quale potrebbe essere il futuro della sinergia uomo-tecnologia in questo campo?

Proseguiremo gli studi e le iniziative progettuali nell'ambito della presa in carico dei cronici lombardi in quanto pensiamo che i benefici dal punto di vista della governance istituzionale possano essere molteplici: dal riconoscimento e ottimizzazione dei percorsi clinici virtuosi sia in termini di efficacia che di efficienza, all'identificazione dei percorsi che con più alta probabilità potrebbero comportare l'aggravarsi della cronicità in essere, includendo la previsione del relativo carico sull'annualità successiva; dal monitoraggio dei percorsi terapeutici

impostati dai gestori a valle della presa in carico, verificando la coerenza del percorso con linee guida e misurandone la variabilità, al monitoraggio del fenomeno di migrazione nel tempo tra i diversi profili di spesa.

L'identificazione di gruppi omogenei di pazienti permette infatti di innestare logiche di valutazione degli esiti e di finanziamento a pacchetto, 'bundle payments'.

Un processo che nel tempo potrebbe anche condurre a nuovi modelli di assistenza sanitaria. A tal proposito mi viene in mente il provocatorio intervento del professor Matthias Mullenbeck, esperto tedesco di immuno-oncologia, in una TED Conference dal titolo "What if we paid doctors to keep people healthy?".

Un giorno il compito dei medici e dei professionisti sanitari potrebbe non essere più quello di dover curare i pazienti affetti da patologie, ma di mantenere in salute i cittadini. È una provocazione, ma probabilmente il futuro è più vicino di quanto ce lo aspettiamo.