

I *BIG DATA* IN ITALIA. LA REALTÀ ATTUALE
E LE IDEE PER IL PROSSIMO FUTURO

1. *Big data e supercalcolo*

Nell'era della trasformazione digitale, due tecnologie, i *big data* e l'intelligenza artificiale (IA), emergono come potenti catalizzatori di innovazione in diversi settori, dalla ricerca, all'industria, all'amministrazione pubblica, alla formazione. Il termine *big data* descrive gli enormi volumi di dati strutturati e non strutturati generati da individui, organizzazioni e macchine, mentre l'IA si riferisce allo sviluppo di sistemi intelligenti, in grado di eseguire automaticamente operazioni e inferenze sofisticate che tradizionalmente richiedevano competenze professionali specialistiche.

La sinergia fra le potenzialità dei *big data* e dell'IA può migliorare i processi decisionali e guidare l'innovazione o il cambiamento socioeconomico in modi mai immaginati prima, tuttavia, la complessità di questa intersezione necessita, per funzionare, di potenti infrastrutture di supercalcolo e di rete, oltre a una decisa spinta all'innovazione tecnologica da parte del legislatore e dei principali attori pubblici ed economici.

I *big data* si riferiscono a insiemi di dati estremamente vasti e spesso poco strutturati, che richiedono approcci specifici per essere sistematizzati, gestiti, elaborati e analizzati in modo efficiente e utile.

Similmente, l'IA, anche riferibile come informatica cognitiva, riguarda lo sviluppo di sistemi software avanzati, tra cui le reti neurali, che possiedono il notevole potenziale di abilitare processi logici, di ragionamento e decisionali in infiniti campi applicativi. L'IA possiede, inoltre, la capacità di essere agile, di apprendere e di adattarsi, anche utilizzando tecniche come la percezione visiva, il riconoscimento delle

Questo capitolo è di Francesco Ubertini e Alessandra Poggiani.

emozioni e la traduzione linguistica. Diversi modelli di calcolo cognitivo abilitati dal *machine learning* (ML) possono essere utilizzati per estrarre informazioni significative dai *big data*, superando grandemente per efficienza i metodi tradizionali.

Tuttavia, a causa delle loro enormi dimensioni e della scarsa organizzazione strutturale, i *big data* non possono essere gestiti con i tradizionali metodi di elaborazione abilitati da data center standard e hanno, invece, bisogno di grandi capacità di calcolo. Lo stesso vale anche per i processi di apprendimento delle applicazioni di IA. Si stima che la produzione globale di dati nel 2025 raggiungerà i 175 zettabytes¹. Un numero enorme: basti pensare che se immagazzinassimo tutti questi dati in tablet da 512 GB di memoria e impilassimo i tablet uno sopra l'altro, realizzeremo una torre alta 5 volte la distanza tra la Terra e la Luna!

Per dirlo in maniera più semplice, i *big data* sono paragonabili a un vastissimo deposito di libri, ma quale sarebbe l'utilità di avere accesso a un deposito così sterminato, senza nessun archivista in grado di trasformare il deposito in una biblioteca per riconoscere e trovare velocemente il libro che cerchiamo? Il supercalcolo è, dunque, lo strumento in grado di analizzare le enormi quantità di dati di cui disponiamo con rapidità e precisione, rendendo i *big data* davvero utili e trasformandoli da un indefinito e infinito bacino di dati grezzi in una serie di informazioni sistematizzate, comprensibili e azionabili secondo uno scopo predefinito. Il supercalcolo è il sistema che rende possibile e veloce l'analisi e la sistematizzazione rapida dei *big data*, permettendo di estrarre interrelazioni e informazioni strategiche da enormi quantità di dati grezzi e disorganizzati, per abilitare decisioni intelligenti e informate.

La straordinaria potenza computazionale dei sistemi HPC (*High Performance Computing*) consente di eseguire rapidamente analisi complesse su dataset di dimensioni enor-

¹ Secondo gli obiettivi del *Digital Decade*, programma strategico della Commissione Europea, che fissa traguardi e obiettivi concreti per il 2030: https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/europes-digital-decade-digital-targets-2030_en.

mi, ottimizzando algoritmi di analisi dei dati e migliorando l'efficienza e l'accuratezza delle analisi, oltre a rendere più rapidi e sostenibili i processi di *ingestion* e apprendimento necessari alle applicazioni di IA.

In sintesi, il supercalcolo fornisce la potenza computazionale necessaria per eseguire analisi avanzate su grandi dataset, ed è l'interdipendenza fra supercalcolo e *big data* che è fondamentale per affrontare le sfide complesse che ci troviamo davanti in ambito scientifico e tecnologico nell'era della trasformazione digitale, oltre che per rendere sempre più «concreta», percorribile e sostenibile la frontiera dell'intelligenza artificiale.

2. *Il supercalcolo: definizione, storia e campi di applicazione dei supercalcolatori in Italia*

Per dare una definizione di supercalcolatore possiamo usare una metafora facilmente comprensibile: un supercomputer sta a un comune computer come un veicolo di Formula Uno sta alle automobili che guidiamo nelle nostre città e sulle nostre autostrade. I supercalcolatori sono infatti macchine uniche, progettate e realizzate *ad hoc* per un committente, che contribuisce alla loro realizzazione.

In quanto prototipi, possono dare prova delle proprie prestazioni soltanto una volta accesi, e richiedono tempo per l'affinamento delle loro potenzialità e per funzionare in maniera ottimale.

I supercomputer, sistemi veloci ad altissime prestazioni anche detti HPC, sono sistemi di calcolo con risorse dislocate e operanti in parallelo, in grado di raggiungere capacità di elaborazione estremamente potenti e capaci di svolgere operazioni molto complesse a elevatissime prestazioni, che richiedono l'analisi di grandissime quantità di dati in tempi rapidissimi.

Oltre all'ordine di grandezza della propria potenza di calcolo, i supercomputer si distinguono anche per la loro architettura. Infatti, sfruttano al massimo e con efficacia le prestazioni di ogni computer nel sistema, ora comunemen-

te misurate in ExaFlops², riuscendo a svolgere più di un trilione (un miliardo di miliardi) di operazioni al secondo, utilizzando l'elaborazione parallela invece di quella seriale. Dunque, i supercomputer non fanno una cosa alla volta, ma svolgono molte attività insieme: tramite più processori, suddividono i problemi in parti e lavorano su queste componenti in parallelo e contemporaneamente.

Un supercalcolatore resta «super» per un periodo limitato, circa cinque anni, dopo i quali diventa un calcolatore normale, considerando la velocità di tasso di innovazione del settore tecnologico.

L'Italia ha una tradizione importante legata al supercalcolo, fin da quando sul finire degli anni Sessanta, quattro rettori di altrettante università (Alma Mater Studiorum di Bologna, Università di Firenze, Università di Padova e Istituto universitario di Economia e commercio e di Lingue e letterature straniere di Venezia) decisero di costituire il Consorzio interuniversitario per la gestione del Centro di calcolo elettronico dell'Italia nord-orientale, dando vita alla prima fase del CINECA. L'idea di base era quella di unire le forze per realizzare il primo supercalcolatore italiano, ai tempi tecnologia di frontiera, a favore della comunità scientifica. Ciò che ha poi portato nel tempo il CINECA a diventare uno dei maggiori centri di supercalcolo in Europa è stato, dunque, lo sguardo lungimirante di quattro università, che – come spesso accade nel campo della ricerca – hanno permesso di raccogliere frutti allora inimmaginabili.

Da quel primo passo, l'Italia ha avuto un discreto numero di supercomputer. Basti pensare che Leonardo, l'ultimo nato e attualmente in funzione al Tecnopolo di Bologna sotto la gestione di CINECA, è il ventesimo sistema di supercalcolo del nostro paese³.

Tuttavia, Leonardo non è solamente «italiano». Da quando l'Europa ha deciso di dotarsi di una strategia e di una

² In grado di calcolare almeno 10^{18} operazioni (moltiplicazioni e/o addizioni) a doppia precisione (64 bit) IEEE 754 al secondo.

³ <https://www.aicanet.it/storia-informatica/calcolo-scientifico-in-italia/cineca>.

struttura sul supercalcolo nel 2018, istituendo la EuroHPC JU, acronimo di *The European High Performance Computing Joint Undertaking*, i maggiori centri di supercalcolo dell'EU sono diventati parte di un sistema federato, cofinanziato al 50% con risorse di tutti i paesi membri dell'Unione, e per l'altro 50% dai paesi ospitanti.

Col trascorrere del tempo, intanto, è aumentata in maniera esponenziale la quantità di campi di applicazione dei supercalcolatori. Nati per sostenere la ricerca scientifica, ancora oggi i supercomputer sono utilizzati, ad esempio, per analizzare i dati emersi dagli esperimenti dei fisici, per facilitare le osservazioni degli astrofisici che studiano le galassie, o in aiuto ai test che eseguono i fisici sulla Terra solida. Non solo nel campo della fisica, i supercalcolatori hanno un ruolo chiave anche per quanto riguarda la meteorologia e le previsioni climatiche. Ad esempio, il sistema di supercalcolo del CINECA aveva previsto l'entità della recente alluvione in Emilia-Romagna, a maggio 2023, e i dati erano stati forniti alla Protezione civile per diramare l'allerta.

Oltre alla ricerca scientifica, i supercomputer hanno un vasto impiego anche nel settore industriale. Il supercalcolo è fondamentale, ad esempio, per individuare i giacimenti di gas e da molti anni, infatti, le compagnie energetiche utilizzano questi strumenti per anticipare le prospezioni, attraverso studi geologici e fluidodinamici, velocizzando l'individuazione e la messa in produzione di nuovi giacimenti con interessanti casi di successo, a partire da ENI⁴.

Da oltre vent'anni, poi, il supercalcolo viene utilizzato nell'industria automobilistica. Un tempo, la progettazione di nuove auto veniva affinata attraverso i cosiddetti «crash test», che si rivelavano estremamente dispendiosi. Occorreva realizzare un prototipo, sottoporlo a test, e poi verificare le funzionalità da correggere, in maniera ricorsiva e con tempi lunghi e, spesso, incerti. Oggi i supercalcolatori permettono di poter eseguire in maniera affidabile crash test digitali,

⁴ <https://www.eni.com/it-IT/azioni/tecnologie-transizione-energetica/supercalcolo-intelligenza-artificiale/supercomputer.html>.

simulando pienamente quelli veri. In questo modo, il lungo e dispendioso ciclo di correzione descritto sopra, può essere realizzato con l'ausilio di prototipi virtuali. Solo alla fine del percorso di messa a punto viene creata l'auto ed eseguito il crash test fisico.

Un altro ambito in cui i supercalcolatori trovano ampio utilizzo è quello della medicina e in generale delle *life sciences*, ad esempio per quanto riguarda lo sviluppo di nuovi farmaci. Durante la pandemia Covid, i supercalcolatori di tutto il mondo sono stati utilizzati per velocizzare la scoperta di molecole per realizzare medicinali in grado di contrastare la malattia. L'ampissimo database delle molecole conosciute e delle loro proprietà è stato incrociato con gli attributi del virus, ottenendo in pochi secondi tutte le combinazioni possibili, e restringendo enormemente il numero di molecole adatte alla creazione di farmaci efficaci.

Inoltre, le risorse di supercalcolo sono oramai cruciali allo sviluppo della medicina di precisione, determinato dal recente cambio di scala raggiunto nell'acquisizione di dati biologici relativi ai singoli individui, grazie alle cosiddette discipline omiche che, utilizzando sofisticate tecnologie di analisi, consentono la produzione di elevate quantità di dati utili per la descrizione e l'interpretazione del sistema biologico, fra cui, innanzitutto, la genomica, oltre alla proteomica, alla matabolomica e alla trascrittomica.

Nel settore della genetica, i sequenziatori che ricostruiscono la sequenza del genoma umano hanno necessità di essere coadiuvati da calcolo ad alte prestazioni, in grado di processare i dati raccolti, per fornire dati effettivamente utilizzabili.

I modelli predittivi elaborati da algoritmi abilitati dal supercalcolo, possono anche dare vita ai cosiddetti «gemelli digitali»: rappresentazioni virtuali di un oggetto o di un sistema, collegate a esso per tutto il ciclo di vita. Infatti, il «gemello digitale» viene aggiornato in tempo reale dai dati raccolti dai sensori collegati all'oggetto fisico e usa programmi di simulazione e l'apprendimento automatico per elaborare modelli predittivi delle prestazioni future e delle reazioni dell'oggetto a determinate condizioni, che vengono

restituite alla controparte fisica come supporto alle decisioni. In termini più semplici, il gemello digitale è un modello virtuale altamente complesso, che è l'esatta replica del suo corrispettivo fisico. Le potenzialità e i campi di applicazione del *digital twinning* sono numerosi. A oggi i gemelli digitali vengono utilizzati principalmente nell'industria, ma la tecnologia si sta affermando anche nel settore pubblico, soprattutto per le amministrazioni municipali. Sono molti gli esempi di gemelli digitali di città, anche in Europa⁵.

Ma la scala è più ampia. Un visionario programma dell'Unione Europea (a cui CINECA partecipa), intitolato *Destination Earth*⁶ prevede di arrivare al 2030 con un gemello digitale dell'intera Terra, cioè di tutti i fenomeni fisici che avvengono sul nostro pianeta. La logica del gemello digitale non è però solo quella di fornire delle previsioni, ma anche di orientare ed eventualmente correggere i processi decisionali. Se utilizzato in maniera interattiva, cambiando le condizioni, immettendo nuove variabili, è possibile anticipare l'effetto delle scelte ed eventualmente modificarle. Pensiamo, ad esempio, ai cambiamenti climatici: il gemello digitale della Terra potrà essere utilissimo nello studio delle azioni più efficaci per prevedere e contrastare i disastri naturali.

Il supercomputer Leonardo, inoltre, gioca un ruolo chiave nell'allenamento dell'intelligenza artificiale generativa e nella creazione di LLM (*Large Language Models*). Il sistema di supercalcolo è già stato utilizzato per il training di modelli europei *open* e sta lavorando per allenare modelli in lingua italiana.

I numeri parlano chiaro: per allenare ChatGPT-4 ci sono voluti alcuni mesi, per un costo totale stimabile in oltre 100 milioni di dollari. Leonardo avrebbe potuto portare a termine lo stesso addestramento in 120 giorni.

⁵ <https://digital-strategy.ec.europa.eu/en/library/local-digital-twins-forging-cities-tomorrow>.

⁶ <https://destination-earth.eu/>.

3. *La traiettoria italiana nella strategia europea*

Forti dell'esperienza decennale maturata in ambito di supercalcolo, quando su iniziativa della Regione Emilia-Romagna il progetto di rigenerazione urbana dell'area della Manifattura Tabacchi di Bologna diede vita all'ecosistema Tecnopolo come luogo di innovazione per *big data*, supercalcolo e trasformazione digitale, il CINECA e l'INFN hanno messo a sistema le proprie competenze e parte delle proprie infrastrutture. Quasi contemporaneamente, nel 2017, l'Italia ha candidato Bologna per portare al Tecnopolo il data center del Centro meteorologico europeo (ECMWF), di fatto promuovendo quell'area a hub internazionale delle nuove tecnologie di supercalcolo.

Poco dopo, l'Europa ha deciso di dotarsi di una rete di supercalcolatori di classe pre-exascale e, sull'onda dell'insediamento del data center del Centro meteorologico europeo, l'Italia si è candidata per ospitarne uno, sempre nell'area del Tecnopolo di Bologna. Questo supercalcolatore, Leonardo, è entrato in funzione nel novembre del 2022 classificandosi allora come il quarto supercomputer più potente al mondo⁷. La potenza di un supercalcolatore si misura in numero di operazioni al secondo: Leonardo ne fa 250 milioni di miliardi, meritandosi così la definizione pre-exascale.

Per fornire un'idea della sua velocità, possiamo immaginare che ogni singolo abitante della terra compia un'operazione algebrica al secondo, per 365 giorni: questo, in un anno, sviluppa una potenza di calcolo che Leonardo è in grado di eseguire in un solo secondo.

Il sistema, che è costato 250 milioni di euro, è stato trasportato al Tecnopolo con l'ausilio di 30 tir, e contiene al suo interno 14.000 GPU di architettura Ampere, che CINECA ha co-progettato con NVIDIA. Le unità di calcolo sono tutte collegate tra loro attraverso una fibra: unendo tutte le fibre che collegano i nodi di calcolo si coprirebbe una distanza di circa 160 km. La cosa particolarmente interessante di Leonardo è soprattutto la sua tecnologia: c'è

⁷ <https://www.top500.org>.

molto più software disponibile rispetto alle altre macchine più potenti del mondo. Quindi – per riprendere la metafora della Formula Uno – è vero che non è la prima auto al mondo per velocità di giro, ma è quella da tenere d’occhio quanto a possibilità di vincere il Gran Premio.

Le applicazioni basate sull’IA, ad esempio, possono sfruttare GPU all’avanguardia, *core tensors* dedicati e un’architettura di sistema progettata per supportare carichi di lavoro legati all’I/O, grazie anche alla funzione NVIDIA GPUDirect RDMA e allo *storage fast tier*.

Se, come abbiamo detto, la potenza di calcolo è attualmente pari a 250 PFlops, l’obiettivo per il 2026 di CINECA è quello di portare Leonardo a un ExaFlops, ovvero raggiungere il miliardo di miliardi di calcoli al secondo.

Il sistema Leonardo, gestito da CINECA, è utilizzabile gratuitamente da soggetti italiani ed europei per attività di ricerca e innovazione di tipo *open*, attraverso modalità di accesso per merito. Ciclicamente vengono pubblicati avvisi pubblici rivolti a soggetti pubblici e privati, che prevedono una valutazione scientifica e tecnica, una stima delle risorse di calcolo necessarie e del personale tecnico da destinare allo sviluppo dell’attività, in modo da assegnare risorse e tempi.

Tuttavia, come abbiamo detto, un supercomputer rimane tale per un massimo di cinque anni. È naturale, dunque, che si cominci a lavorare a un «dopo Leonardo». Già nel 2024 il sistema verrà aggiornato con un sistema integrativo dal nome «Lisa» che, auspicabilmente, permetterà al supercalcolatore di restare nella top ten mondiale.

Nella seconda parte dell’anno, invece, verrà installato al Tecnopolo un computer quantistico. L’Europa ha deciso infatti di dotarsi di una prima rete di sei computer quantistici e l’Italia, con CINECA, se ne è aggiudicato uno⁸. Saranno macchine allo stato dell’arte, ognuna con una tecnologia differente: quella che sarà installata a Bologna lavorerà insieme a Leonardo a servizio della comunità scientifica.

⁸ https://eurohpc-ju.europa.eu/selection-six-sites-host-first-european-quantum-computers-2022-10-04_en?prefLang=it.

Nel prossimo futuro, già a partire dal 2026, la sfida è di far evolvere ulteriormente Leonardo, con un'architettura sempre più orientata ad applicazioni IA, e superare un ExaFlops di potenza.

4. *Il futuro dell'Europa nella sfida della trasformazione digitale*

Attualmente tre dei più potenti supercalcolatori del mondo sono nell'Unione Europea, eppure dentro ai supercalcolatori esistenti, c'è pochissima tecnologia europea. Il nostro continente si trova pericolosamente indietro nell'industria della componentistica ad alto tasso di innovazione tecnologica, scontando un deficit di politica industriale che non permette, oggi, all'Europa di essere indipendente o, quantomeno, non dipendente da altri paesi.

Il ritardo europeo nel settore è un tema molto rilevante e oramai inserito a pieno titolo nell'agenda del legislatore europeo. L'Unione Europea nel 2022 ha finalmente promosso il *Chips Act*⁹, la legge europea sui semiconduttori, che prevede lo stanziamento di 43 miliardi di euro per la creazione di una filiera europea di design e produzione dei chip.

Questo ritardo ha forti ripercussioni su tutta l'industria europea. Di nuovo, il settore automobilistico fornisce un rilevante esempio di quello che sta succedendo. L'automobile sta cambiando rapidamente forma: non ha più bisogno del volante, si guida da sola, è sempre più in rete, ha sempre più bisogno di capacità di calcolo e software per gestire sia la guida autonoma che gli accessori evoluti. La macchina del futuro sarà praticamente un computer su quattro ruote. Per gestire la guida autonoma servono chip avanzati, paradossalmente molto simili a quelli usati nei supercomputer. NVIDIA, corporation statunitense, nata come società per il *gaming* e ora specializzata nella tecnologia di supercalcolo, è l'impresa più avanzata nello sviluppo di questi nodi e del software per la gestione delle auto del futuro.

⁹ https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-chips-act_it.

In pratica, il settore dell'*automotive*, che da sempre ha importanti insediamenti in Europa, sta cambiando velocemente e lo strato più alto del valore aggiunto sarà nelle tecnologie digitali. Quello strato è al momento ideato e realizzato esclusivamente fuori dall'Europa, in un settore con politiche industriali molto aggressive e competitive, di fatto facendo rischiare al nostro continente di perdere un comparto produttivo molto importante per l'economia e l'occupazione.

Lo sviluppo in Europa di un ecosistema di supercalcolo competitivo e di una capacità di calcolo quantistico integrata di livello mondiale, potrà aiutare a colmare alcuni ritardi e a far recuperare all'UE una posizione di leadership nell'economia dei dati, rafforzando l'autonomia dell'Europa in alcuni settori abilitati dal supercalcolo, come quello dell'IA.

5. *Il ruolo della formazione terziaria nella sfida tecnologica*

Perché il «salto» europeo avvenga, la condizione è garantire l'evoluzione dell'industria del software e della componentistica per il supercalcolo, in maniera sinergica e sincrona.

Fra le sfide principali, c'è quella di investire su una nuova generazione di professionalità e di ricercatori che possa rendere tutto questo possibile. L'Unione Europea si sta impegnando in maniera significativa per conseguire questi ambiziosi obiettivi.

L'evoluzione tecnologica che mette l'analisi delle informazioni, i *big data* e l'IA al centro dello sviluppo e dell'innovazione, ha portato anche a riconsiderare la necessità di coniugare competenze scientifiche e tecniche con competenze umanistiche. I più recenti programmi dell'Unione mirano a promuovere un approccio STEAM¹⁰, che indica un insieme di approcci multidisciplinari all'istruzione per eliminare le barriere tradizionali tra materie e discipline tecnologiche e scientifiche e quelle umanistiche o delle scienze sociali, favorendo il trasferimento di conoscenze tra settori.

¹⁰ <https://steamit.eun.org/>.

Le università italiane debbono sapere rispondere a questa sfida non solo con proposte didattiche innovative, ma anche con nuovi servizi agli studenti e al personale docente e non docente, e con metodologie d'insegnamento in linea con il contesto tecnologico. Si sono evolute le competenze richieste, sono mutate profondamente le professionalità più ricercate, è aumentata in maniera esponenziale la domanda nel settore tecnologico, ma, ancora, l'istruzione universitaria italiana non sembra rispondere in maniera efficace, se è vero, come si evince da una recente ricerca sul campo¹¹, che il disallineamento fra domanda e offerta nel mercato del lavoro è sempre più preoccupante (solo per citare un esempio, nel 2022 gli annunci di lavoro per sviluppatori software sono stati 171.048, a fronte di 8.998 laureati nel 2021 con competenze adatte a questo impiego).

Le università devono essere il soggetto che anticipa e guida i cambiamenti e, di fronte alle grandi sfide poste dai cambiamenti tecnologici e sociali, devono trovare, nel contesto in cui operano, degli elementi caratterizzanti in grado di assicurare loro un ruolo cruciale e ragione di esistere.

Poiché nel mondo attuale la dimensione digitale è parte integrante dell'identità degli studenti, l'università non può che essere ibrida: la digitalizzazione dei processi di apprendimento e condivisione dei saperi è un elemento imprescindibile, seppure non sostitutivo dei percorsi di studio «tradizionali».

Nella digitalizzazione della didattica può offrire un grande aiuto l'IA generativa, in grado di segnare un punto di svolta fondamentale anche in questo campo specifico. Paradossalmente, l'unica cosa che manca all'intelligenza artificiale è proprio l'intelligenza: questo strumento non può prescindere dal controllo umano, e per essere allenato, guidato e utilizzato al meglio richiede competenze specifiche e grande preparazione. La formazione diventa (ritorna!)

¹¹ *Mercato delle competenze e dottorati di ricerca. Il difficile dialogo tra università e mercato del lavoro*, survey Tomorrow Education, Talents Venture, maggio 2023: https://www.talentsventure.com/wp-content/uploads/2023/05/D2.D3-2023-Mercato-delle-competenze-e-dottorati-di-ricerca_webinar.pdf.

dunque cruciale per governare l'intelligenza artificiale e non subirla come meri consumatori.

L'Italia è in ritardo su alcuni segmenti della formazione, ad esempio sull'istruzione di terzo livello a ciclo breve (ITS), e soprattutto sulle lauree triennali, in cui le percentuali nazionali sono staccate di 10-15 punti rispetto alle altre economie avanzate. Il ritardo è sia nel numero di giovani diplomati che si iscrivono all'università o a percorsi biennali ITS, sia in quelli che una volta iscritti riescono poi a conseguire il titolo. In altre parole, nel nostro paese ci sono troppi giovani che si fermano al diploma anziché proseguire gli studi, e troppi che cominciano il percorso per poi abbandonarlo prima della conclusione.

Per contrastare questa tendenza negativa e rispondere alla domanda di formazione sui temi della digitalizzazione, la didattica deve farsi più flessibile, dando spazio a competenze «trasversali» e favorendo la contaminazione tra aree umanistiche e tecniche.

Nell'ottica di ridurre al minimo il divario tra domanda e offerta di cui parlavamo poco fa, diventa fondamentale ridisegnare percorsi di laurea triennale professionalizzanti, che siano pensati in sinergia con il mondo del lavoro e delle imprese.

6. *Trasformazione digitale e big data per innovare le università*

La sfida delle università non si limita, tuttavia, al disegno di nuovi percorsi formativi, più capaci di rispondere alle esigenze del contesto mutato. Sono anche le forme, non solo il contenuto della formazione, a dover cambiare.

L'utilizzo proficuo e specifico dei *big data* e di applicazioni di IA ha grandi potenzialità trasformative. L'analisi dei dati può essere senz'altro uno strumento utile a comprendere meglio i comportamenti degli studenti (ad esempio, aiutando la comprensione dell'abbandono scolastico) e a facilitare il lavoro del personale docente e non docente.

In questo senso, il cosiddetto *data mining* può fornire un aiuto più che concreto. Si tratta di un processo di

esplorazione e analisi di un grande quantitativo di dati, con l'obiettivo di individuare eventuali regolarità, o estrapolare atteggiamenti ricorsivi. Grazie al *data mining* si possono ottenere informazioni da un database e organizzarle in pattern, mettendo così in evidenza associazioni e correlazioni.

D'altra parte l'IA è già utilizzata sperimentalmente (e può essere molto sviluppata) per sostenere gli studenti in percorsi di apprendimento differenziato o individualizzato, oppure per superare alcune disabilità, ad esempio con sintetizzatori vocali di testi o tutor per studenti con deficit cognitivi, solo per citare alcuni esempi.

Attraverso l'IA e processi di *machine learning* si possono sviluppare ambienti di *social annotation* che consentono di coinvolgere gli studenti in discussioni strutturate finalizzate ad affrontare gli aspetti più critici o complessi delle lezioni per promuovere il miglioramento del percorso di apprendimento.

È sempre più necessario, quindi, che ricercatori, docenti e strutture amministrative degli atenei comprendano le potenzialità dell'IA e dell'utilizzo dei dati per avvalersi delle tecnologie in modo positivo e sfruttarne appieno le potenzialità.

Nell'ottobre 2022, per contribuire ad affrontare queste sfide, la Commissione ha pubblicato le prime linee guida sull'uso dell'IA e dei dati nell'insegnamento e nell'apprendimento¹² che affrontano, sotto il profilo etico, le modalità di utilizzo dell'IA per aiutare docenti e educatori a comprendere il potenziale che le applicazioni dell'IA e dell'uso dei dati possono avere nell'istruzione e sensibilizzarli sui possibili rischi. L'auspicio è che sia un primo passo per accompagnare l'intero settore dell'istruzione superiore a una maggiore consapevolezza affinché la tecnologia sia un alleato nella costruzione del futuro delle nuove generazioni.

¹² European Commission, *Ethical Guidelines on the Use of Artificial Intelligence (AI) and Data in Teaching and Learning for Educators*, Bruxelles, 25 ottobre 2022, <https://op.europa.eu/en/publication-detail/-/publication/d81a0d54-5348-11ed-92ed-01aa75ed71a1/language-en>.