

Position paper

AI e etica

Il ruolo dell'etica nella strategia di NTT DATA
per lo sviluppo di una AI sostenibile

Indice dei contenuti

01	AI e etica: la nuova frontiera della responsabilità digitale	1
	Bilanciare innovazione e principi morali	
02	AI Act – Il regolamento europeo	2
	I rischi di un uso improprio dell'intelligenza artificiale	
03	Linguistica digitale e comportamento umano	3
	Governare l'interazione uomo-macchina nell'era dell'intelligenza artificiale conversazionale e generativa	
04	L'approccio di NTT DATA	4
	Team interdisciplinari per un uso etico dell'intelligenza artificiale	
05	Dalla teoria alla pratica	6
	Alcuni use case relativi all'applicazione del digital linguistics & human behaviour di NTT DATA	

AI e etica: la nuova frontiera della responsabilità digitale

Bilanciare innovazione e principi morali

L'intelligenza artificiale è una tecnologia che sta attirando l'attenzione di aziende ed istituzioni italiane, destinata - secondo le stime degli analisti - a crescere in modo importante nei prossimi anni, passando da un valore di spesa di 960 milioni di euro nel 2024 a quasi due miliardi di euro nel 2027 in Italia (fonte: Netconsulting cube, 2024).

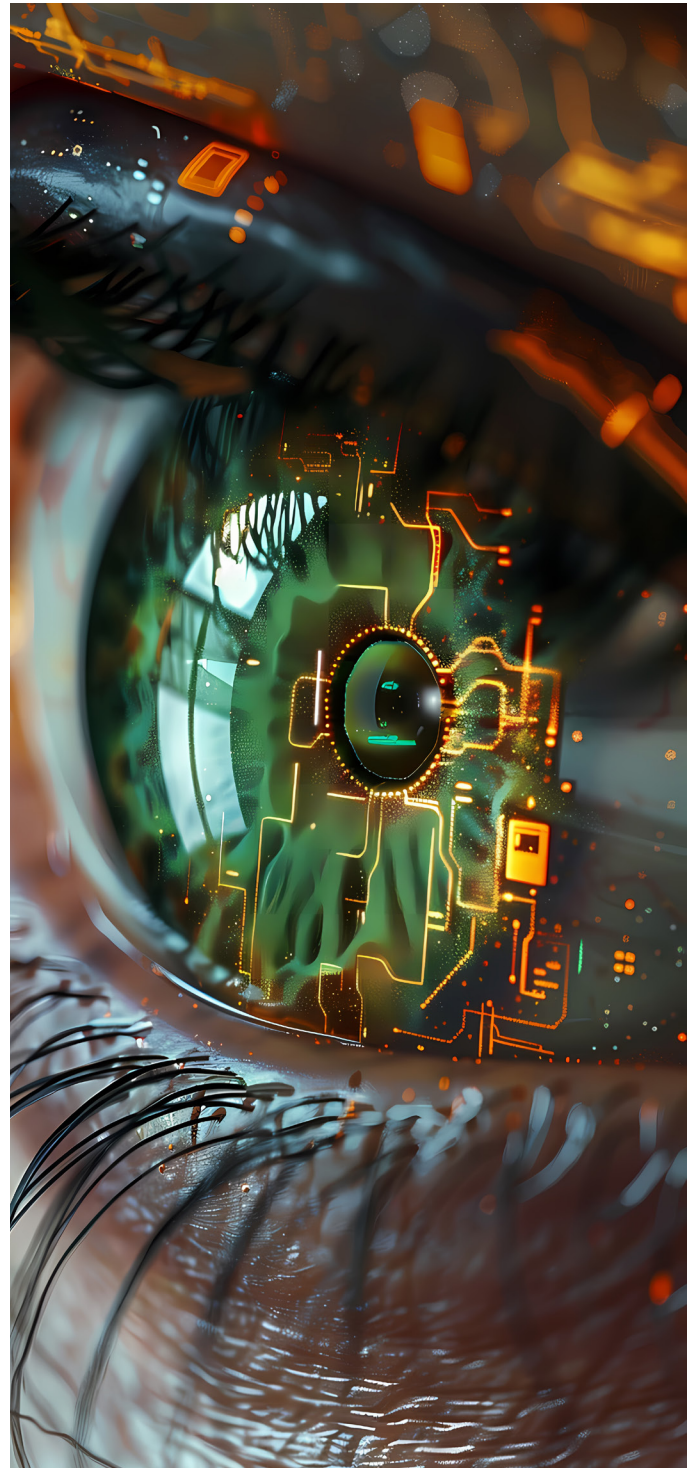
Come tutte le tecnologie dirompenti che hanno un impatto non solo sulla vita aziendale ma anche sulla vita "reale", **l'intelligenza artificiale impone riflessioni** che abbracciano molteplici aspetti. Uno su tutti l'etica, il rapporto tra l'intelligenza artificiale e i valori che la governano, grazie a un approccio collettivo che coinvolge profili di estrazione tecnologica, derivazione giuridica, sociologica, umanistica e linguistica.

In questo contesto, **l'etica** deve essere considerata come un insieme di principi normativi e valori universali in grado di guidare lo sviluppo di soluzioni di intelligenza artificiale per supportare un progresso tecnologico sostenibile, capace di rispettare i diritti dell'uomo e promuovere il bene comunitario. Tale questione è di fondamentale importanza, considerato che si stanno già prefigurando scenari in cui l'intelligenza artificiale verrà applicata a settori - quali la sanità e i sistemi giudiziari - dove le considerazioni etiche divengono cruciali ed imprescindibili.

Grazie all'AI tutti noi siamo messi nelle condizioni di poter svolgere attività in modo più rapido e preciso, ma siamo anche maggiormente esposti al rischio di minare **norme e principi** di riferimento, come la privacy, la cybersicurezza, l'affidabilità di dati e informazioni, la responsabilità civile e penale.

Se lo scopo dell'intelligenza artificiale è quello di mettere le persone, e non i sistemi, nella condizione di prendere decisioni informate in tempi più veloci (basandosi su dati più accessibili rispetto al passato), allora è importante assicurarsi che l'etica sia rispettata in tutte le fasi della catena del valore dell'AI, dall'ideazione e progettazione, alla formazione sui large language model (LLM), alle competenze delle figure preposte al suo utilizzo.

Questo livello di attenzione è necessario per gestire il rapporto con questa tecnologia, i criteri e i principi che guidano il suo sviluppo. Ma la **fiducia** nell'AI non dovrà mai essere incondizionata, saranno sempre le persone, e non i sistemi, a prendere decisioni consapevoli.



AI Act – Il regolamento europeo

I rischi di un uso improprio dell'intelligenza artificiale

Nel contesto in cui ci troviamo ora – polarizzato tra chi ha grandi aspettative legate all'AI per aziende e società e chi teme un impatto su posti di lavoro o utilizzi fraudolenti – emerge la necessità di regolamentare sia come l'AI viene istruita e educata sia come viene utilizzata.

Per questo motivo a marzo 2024 il parlamento europeo ha approvato l'**AI Act**. Il primo documento a livello mondiale che mette il controllo delle tecnologie di AI al centro del progresso dell'uomo, della società e dell'economia e che **entrerà in vigore gradualmente**, per arrivare a un'applicazione completa entro la prima metà del 2026.

L'AI Act ha come obiettivo primario quello di proteggere i diritti fondamentali, la democrazia, lo stato di diritto e la sostenibilità ambientale dai sistemi di intelligenza artificiale ad alto rischio, promuovendo contemporaneamente l'innovazione e assicurando all'Europa un ruolo guida nel settore, stabilendo obblighi per l'AI sulla base dei **possibili rischi** e del livello d'impatto.

L'AI Act gioca proprio un ruolo importante nell'indirizzare un utilizzo etico della tecnologia che rispetti **trasparenza, inclusione, responsabilità, imparzialità, affidabilità, sicurezza e privacy**, per garantire uguaglianza di trattamento ed evitare discriminazioni.

Rischio e trasparenza sono due degli elementi su cui vale la pena focalizzare l'attenzione. Da un lato, l'approccio basato sul rischio prevede che l'uso dell'AI in applicazioni più sensibili sia soggetto a regole severe e controlli rigorosi. Per esempio, per i sistemi di AI ad alto rischio vige l'obbligo di mantenere registri d'uso, essere trasparenti e accurati e garantire la sorveglianza umana. Rientrano all'interno di questa casistica l'applicazione dell'intelligenza artificiale a infrastrutture critiche, istruzione e formazione professionale, occupazione, servizi pubblici e privati di base (banche, sanità), gestione delle frontiere, giustizia e processi democratici come le elezioni. Esistono poi divieti specifici per l'uso dell'AI in

attività di polizia predittiva, identificazione biometrica e riconoscimento delle emozioni.

Sono previste comunque alcune **deroghe** a queste restrizioni per gestire alcune situazioni eccezionali, come una minaccia terroristica imminente, la ricerca di vittime di gravi crimini o la persecuzione di reati gravi o abusi. Ad esempio, l'identificazione biometrica (Rbi) potrà essere adottata "in tempo reale" solo in casi specifici, con un uso limitato nel tempo e nello spazio e dopo una autorizzazione giudiziaria o amministrativa.

Eccezioni sono poi previste anche per i modelli utilizzati esclusivamente per scopi di ricerca. L'AI Act a tal proposito prevede la creazione di sandbox regolamentari, dove sperimentare soluzioni di AI in condizioni reali, ma con esenzioni regolamentari per promuovere l'innovazione. Sono previste eccezioni specifiche per le **piccole e medie imprese** per facilitare l'adeguamento alle nuove normative, riconoscendo le loro limitazioni di risorse rispetto ai big player del settore.

Dall'altro lato, le aziende che sviluppano potenti intelligenze artificiali generative devono garantire maggiore trasparenza, fornendo informazioni sulle fonti e sui metodi di addestramento, nel rispetto della legge europea sul diritto d'autore. Inoltre, tutti i contenuti prodotti con l'intelligenza artificiale – immagini, testi e video – devono essere chiaramente etichettati per garantire la trasparenza agli utilizzatori.

L'AI Act, in sintesi, vuole evitare di creare un'algocrazia – letteralmente "dominio degli algoritmi" – a favore dell'**algotetica** – letteralmente "etica dell'algoritmo" – per offrire a tutti gli individui una intelligenza artificiale senza pregiudizi, salvaguardando la dignità umana. L'algotetica – termine creato da Paolo Benanti, francescano a capo del gruppo di lavoro che definisce la strategia italiana sull'AI e membro del Comitato sull'intelligenza artificiale delle Nazioni Unite – è un insieme di principi che guida il funzionamento dell'AI: una sorta di guardrail etico capace di mantenere l'AI all'interno di percorsi accettabili.

Linguistica digitale e comportamento umano

Governare l'interazione uomo-macchina nell'era dell'intelligenza artificiale conversazionale e generativa

In questa fase di diffusione dell'AI è necessario studiare quali siano i problemi etici e i risvolti sociali, politici, economici e organizzativi che derivano dall'uso delle tecnologie, soprattutto ora che l'interazione uomo-macchina avviene anche in forma conversazionale.

La **governance dell'AI** deve andare ben oltre gli aspetti tecnologici e deve coinvolgere differenti discipline, dalla tecnologia al diritto, dalla sociologia alla filosofia affinché le persone (e non le "macchine" o i "sistemi") siano nella condizione di prendere decisioni consapevoli. Per questo soluzioni di AI necessitano di essere portate avanti da gruppi misti, interfunzionali, all'interno dei quali ingegneri e sviluppatori software devono collaborare non solo con i data scientist ma anche con figure umanistiche quali i linguisti, che lavorano sulla componente più "umana" dell'intelligenza artificiale.

È proprio questo l'obiettivo della **Digital Linguistics & Human Behaviour**, struttura in **NTT DATA** che si occupa dei temi di AI e NLP e che si posiziona all'incrocio tra linguistica tradizionale, informatica e scienze sociali. Invece di concentrarsi sull'uso dei computer per svolgere attività linguistiche (come la traduzione automatica o il riconoscimento vocale, sottodomini della linguistica computazionale), le tecnologie di Digital Linguistics & Human Behaviour puntano a migliorare l'interazione uomo-macchina in linguaggio naturale, creando esperienze empatiche, anticipando le esigenze degli utenti

e proponendo soluzioni alternative per facilitare il loro lavoro e la loro vita.

Obiettivo principale è mettere l'AI nella condizione di meglio comprendere il linguaggio naturale dell'uomo, ribaltando l'approccio tradizionale secondo cui è l'uomo a dover imparare il linguaggio dei sistemi.

Le tre aree tecnologiche della Digital Linguistics & Human Behaviour che contribuiscono a migliorare la comunicazione sono: Natural Language Processing (NLP), Internet of Behaviour (IoB) e Digital Human.

Il Natural Language Processing (NLP) è un campo dell'intelligenza artificiale che consente alle macchine di comprendere e rispondere al linguaggio umano, elaborando grandi volumi di dati per estrarre informazioni pertinenti.

L'Internet of Behaviour (IoB) raccoglie e analizza dati comportamentali delle persone attraverso dispositivi connessi a Internet, combinando IoT e psicologia comportamentale per influenzare decisioni e azioni future.

I Digital Human sono rappresentazioni virtuali di esseri umani che utilizzano intelligenza artificiale, modellazione 3D e animazione, per fornire esperienze conversazionali coinvolgenti in settori come intrattenimento, assistenza clienti, formazione e telemedicina.



L'approccio di NTT DATA

Gruppi interdisciplinari per un uso etico dell'intelligenza artificiale

La Digital Linguistics & Human Behaviour di NTT DATA è formata da un **gruppo eterogeneo di persone** con competenze tecnologiche e umanistiche differenti che si prefigge l'obiettivo di facilitare un approccio etico all'utilizzo dell'Intelligenza Artificiale sia internamente e che nei progetti sui clienti.

L'approccio di NTT DATA parte dalla consapevolezza che per affrontare problemi aziendali complessi è necessaria una prospettiva capace di unire **hard e soft skill**, che può sviluppata solo all'interno di gruppi ibridi in cui le competenze devono contaminarsi vicendevolmente.

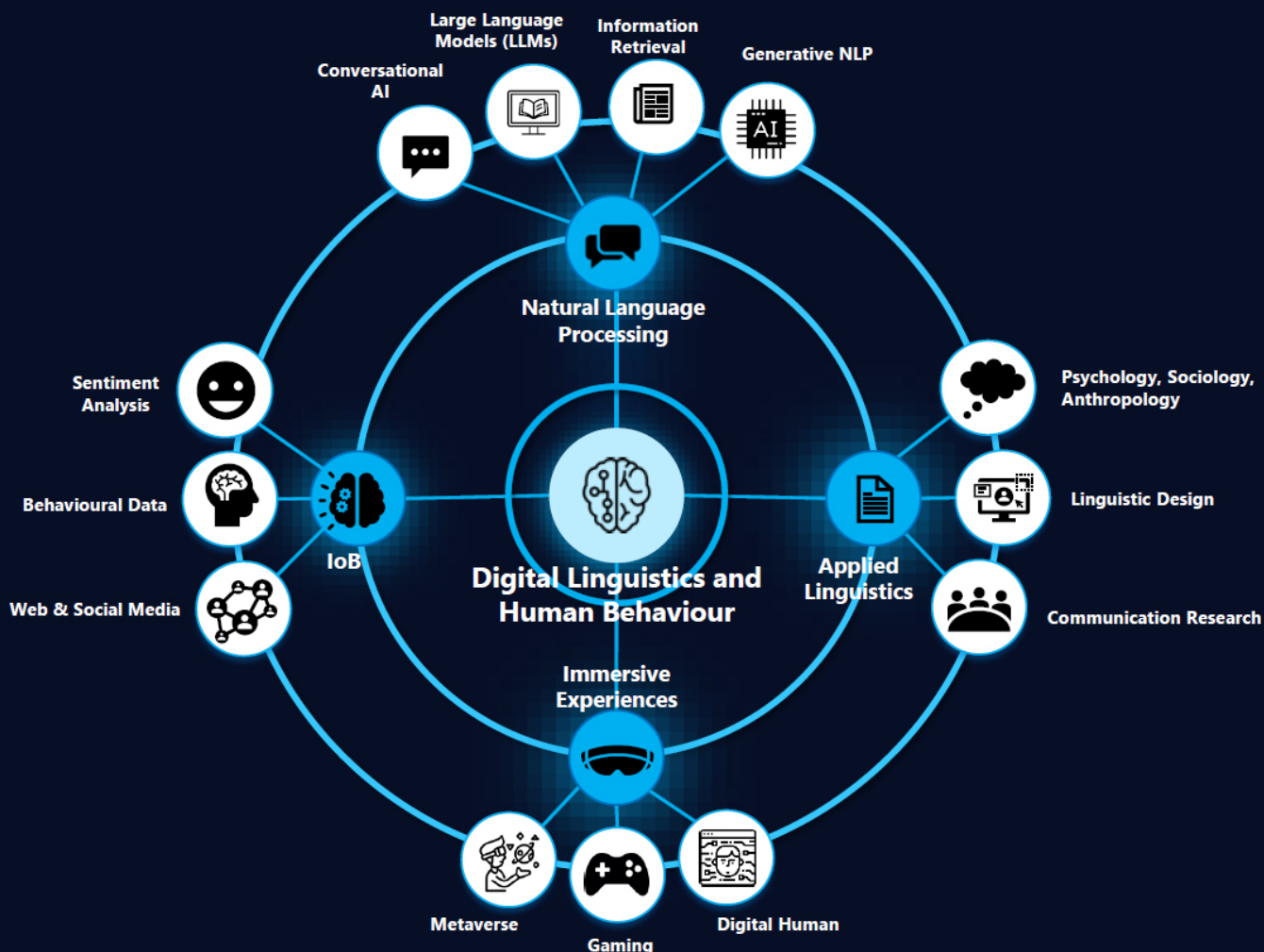
Per NTT DATA le tecnologie di elaborazione del linguaggio naturale (NLP) devono costruire una relazione empatica tra utenti e macchine, al fine di creare esperienze piacevoli, anticipare le esigenze degli utenti e proporre soluzioni alternative per facilitare lavoro e vita.

In questo contesto le tecnologie NLP esaminano grandi volumi

di dati difficili da gestire, estraendo informazioni rilevanti dai documenti. Inoltre, consentono di svolgere compiti comuni come il riconoscimento vocale, la comprensione e la generazione di linguaggio naturale.

Attraverso poi l'**Internet of Behaviour** (IoB) i dati raccolti da diverse fonti – inclusi i dispositivi domestici ed i wearable (come smartwatch, fitness tracker e whistler) – vengono analizzati per individuare schemi che possono essere utilizzati dai gruppi di marketing e vendita per guidare il comportamento degli utenti.

Infine, le **tecnologie Digital Human** sono soluzioni progettate per rendere più empatici gli assistenti virtuali: combinano strategie linguistiche, elementi ritmici delle parole e teorie comportamentali, al fine di definire esperienze conversazionali migliori tra esseri umani e macchine. L'obiettivo è creare assistenti virtuali più umani progettando personalità specifiche, scrivendo dialoghi veritieri e gestendo l'intonazione delle loro voci.



L'approccio del Digital Linguistics & Human Behaviour di NTT DATA prevede un ulteriore elemento: **Applied Linguistics**, una metodologia interdisciplinare di sviluppo che identifica soluzioni a problemi della vita reale legati al linguaggio. È all'interno di questo ambito che emerge in modo preponderante la componente umanistica, che ricorrendo alle competenze di sociologi, antropologi e psicologici cerca di capire l'impatto del linguaggio sul comportamento umano.

L'Applied Linguistics si prefigge un duplice obiettivo. Da un lato sviluppare large language model (LLM) su cui addestrare i modelli di intelligenza artificiale che siano il più possibile liberi da bias di qualunque genere. Al riguardo anche l'UNESCO ha espresso la sua preoccupazione riguardo l'uso improprio di LLM che portano a pregiudizi di genere, sottolineando come gli sforzi per mitigarli, finora compiuti da molte aziende, non siano stati sufficienti.

Dall'altro educare gli utenti ad una formulazione di prompt che non inducano l'intelligenza artificiale a risposte discriminatorie. Per questo motivo sta nascendo la nuova figura professionale dei **prompt engineer**, con una formazione non solo tecnologica, in grado di ottimizzare le richieste da sottoporre all'AI per mitigare bias indesiderati e garantire interazioni più etiche.

Nell'approccio di NTT DATA verso una AI etica, la persona si posiziona al centro perché capace di portare valore e potenziare i sistemi. Per questo motivo i team devono essere misti e composti da ingegneri e linguisti, perché possano essere valorizzati tanto gli aspetti tecnologici quanto gli aspetti umanistici.

Gli attori dell'AI dovrebbero compiere tutti gli sforzi ragionevoli per minimizzare ed evitare di rafforzare o perpetuare applicazioni e risultati discriminatori o pregiudizievoli durante l'intero ciclo di vita del sistema di AI per garantire l'equità di tali sistemi.

Raccomandazione dell'UNESCO
sull'etica dell'Intelligenza Artificiale

Per NTT DATA quindi un **prompt efficace** ha tre differenti elementi:

- **Inizio semplice:** il prompt deve essere essenziale all'inizio, per poi essere via via migliorato con elementi di contesto aggiuntivi. È importante fare versioni successive del prompt e scomporre compiti complessi in sotto-compiti più semplici.
- **Istruzioni chiare:** il prompt deve avere istruzioni chiare, fondamentale usare imperativi.
- **Esempi specifici:** il prompt deve essere il più specifico possibile ed è cruciale fornisca esempi chiari e indicazioni su cosa fare (evitando di dire cosa non fare).



Visione umana del problema

Soluzioni basate su modelli
linguistici e psicologici



Visione tecnologica del problema

NLP generativa, algoritmi
cognitivi e infrastrutture
innovative

Dalla teoria alla pratica

Alcuni use case relativi all'applicazione del Digital Linguistics & Human Behaviour di NTT DATA

G.E.N.T.L.E.

G.E.N.T.L.E. è l'acronimo di **G**ender **N**eutrality **T**ool for **L**anguage **E**valuation ed è uno strumento che NTT DATA Italia ha sviluppato in collaborazione con il Politecnico di Milano per aiutare le aziende ad avere uno stile di comunicazione neutro dal punto di vista del linguaggio, eliminando gli stereotipi di genere e cogliendo i benefici di una forza lavoro diversificata.

Lo strumento si basa su algoritmi di NLP integrati con modelli sintattici, linguistici e semantici che puntano a quantificare i pregiudizi di genere nelle comunicazioni aziendali (come l'androcentrismo, gli stereotipi, l'inclusività); valutare il livello di uguaglianza di genere dell'azienda; fornire approfondimenti sullo stile di comunicazione adottato.

La piattaforma, basandosi sui risultati dell'analisi linguistica, suggerisce modifiche dei testi della comunicazione che rispettino i criteri di inclusion e diversity a cui le aziende vogliono aspirare.

Behavioural Gamification

La Behavioural Gamification integra meccanismi di gioco in ambienti e contesti lavorativi come siti web, learning management systems o intranet per aumentare la partecipazione e migliorare l'efficienza dei sistemi.

In questo contesto sono coinvolti soprattutto linguistica applicata e algoritmi di Internet of Behaviour (IoB), che forniscono una user experience conversazionale per favorire la creazione di giochi di ruolo più immersivi e personalizzati. A queste si aggiungono interazioni e dialoghi più realistici e convincenti anche con i personaggi non giocanti, contenuti formativi partendo dall'analisi del comportamento dei giocatori e suggerimenti per migliorare prestazioni e acquisire nuove competenze attraverso i giochi.

Narrative Medicine

Narrative Medicine è una piattaforma per promuovere la collaborazione tra i diversi specialisti del personale medico, per definire un approccio terapeutico migliore.

La piattaforma agisce aggregando tutte le informazioni sui pazienti provenienti da fonti differenti ed eterogenee quali visite mediche, terapie e narrazione dei pazienti.

Da un lato unisce le feature degli algoritmi di NLP e di linguistica applicata per analizzare automaticamente i dati in linguaggio naturale. Dall'altro unisce algoritmi di IoB con l'intelligenza emotiva per catturare le emozioni reali in modo empatico, comprendere il quadro patologico individuale e fornire utili spunti sia per i pazienti che per gli operatori sanitari e i caregiver.

Il Position Paper “**AI e Etica. Etica e sviluppo di una AI sostenibile nelle strategie di NTT DATA**” è stato curato da **Inno3**, testata giornalistica del **Gruppo NetConsulting cube**, attivo nel settore delle analisi di mercato, della consulenza ICT e dei servizi di comunicazione digitale per vendor e aziende utenti.

Il Position Paper è stato redatto per conto di NTT DATA Italia nel mese di Settembre 2024.

Autori:

Antonio Schiano, Responsabile dell'area Digital Linguistics & Human Behaviour di NTT DATA Italia

Sergio Patano, Practice Leader di Sirmi – Gruppo NetConsulting cube

