Dati, etica e innovazione responsabile nello sviluppo delle tecnologie avanzate

di Antonino Rotolo

11-06-2026

Nuovi paradigmi etici per l'HPC e l'AI Agentica

Il Centro Nazionale ICSC è un'infrastruttura strategica che contribuisce alla sovranità digitale dell'Italia. High Performance Computing, Big Data e Quantum Computing non sono più ambiti di ricerca isolati, ma motori integrati d'innovazione sociale e industriale che alimentano domini vitali come la sanità, l'energia, la pubblica amministrazione e la democrazia digitale. In questo scenario, è fondamentale comprendere che i dati non sono un semplice "combustibile" neutro. Essi costituiscono una vera e propria infrastruttura cognitiva e normativa: il modo in cui scegliamo di raccoglierli, modellarli e ottimizzarli riflette e, al contempo, determina le priorità delle politiche pubbliche, influenzando la distribuzione e protezione di diritti, opportunità e risorse. Come recitava una celebre pubblicità degli anni Novanta, "la potenza è nulla senza il controllo", e per ICSC il controllo non è solo un parametro tecnico, ma un orizzonte etico.

L'accelerazione senza precedenti dell'intelligenza artificiale generativa e l'emergere dei sistemi agentici (Agentic AI) hanno messo in crisi i modelli tradizionali di valutazione d'impatto e di rischio delle tecnologie digitali. Se fino a poco tempo fa i sistemi di AI erano visti come prodotti, strumenti in grado di apprendere ma senza cambiare le proprie capacità in modo strutturale, oggi ci confrontiamo con entità capaci di proattività, scomposizione di obiettivi complessi e apprendimento continuo. In un contesto HPC, dove queste entità possono operare su scala massiva, i benefici potenziali per la società sono immensi, ma i rischi si trasformano da "errori locali" a impatti sistemici. L'attenzione di ICSC per le implicazioni sociali ed etiche dell'HPC nasce proprio per governare questa transizione, elaborando cornici concettuali e metodologiche dove la ricerca non si limiti a essere tecnicamente eccellente, ma sia intrinsecamente responsabile. L'obiettivo strategico di ICSC è superare l'idea di etica come "aggiunta a valle" per approdare a una responsabilità sistemica progettata by design lungo l'intero ciclo di vita dei sistemi.

Oltre i principi: responsabilità sistemica e multidimensionale

Negli ultimi anni, il dibattito globale sull'AI si è cristallizzato attorno a principi fondamentali come la centralità dell'essere umano, la trasparenza e l'equità. Tuttavia, l'esperienza maturata all'interno di ICSC evidenzia i limiti di un approccio puramente "principialistico". Il problema risiede nel fatto che i principi etici non sono gerarchici ex ante; spesso, essi entrano in conflitto diretto. Garantire la massima privacy può significare ridurre la trasparenza algoritmica, così come ottimizzare l'efficienza di un sistema può portare a discriminazioni involontarie contro gruppi vulnerabili. Non è possibile rispettare ogni principio al massimo grado simultaneamente: è necessario gestire consapevolmente i trade-off attraverso quello che i giuristi definiscono "bilanciamento definitorio" (definitional balancing).

Per tradurre questi concetti in pratica, è innanzitutto necessario considerare la responsabilità morale e giuridica. La prima è legata alle funzioni specifiche del ruolo ricoperto dai ricercatori e dagli sviluppatori; la seconda riguarda i doveri imposti dalle norme e dai contratti. Questo passaggio richiede una metodologia robusta di valutazione del rischio che vada oltre il singolo bias evitando di riprodurre o rafforzare pregiudizi e discriminazioni già presenti nella società. Dobbiamo parlare di rischio sistemico, caratterizzato da un'alta velocità di propagazione e da effetti a cascata che

possono minare le fondamenta delle istituzioni democratiche. Un esempio critico è la disinformazione di genere nelle campagne elettorali: in questo caso, non siamo di fronte a un singolo errore del modello, ma a una distorsione strutturale che si auto-alimenta, rendendo difficile la reversibilità degli effetti una volta che il contenuto si è diffuso viralmente. L'integrazione di competenze etico-giuridiche nella progettazione delle pipeline AI/HPC non è quindi un adempimento formale, ma il cuore pulsante di una cultura della ricerca che interroga le finalità del sistema prima ancora di scriverne il codice.

La governance dei dati tra qualità, potere e "black mirrors"

La qualità dei dati è il prerequisito per un'AI affidabile, ma la questione è più profonda della mera pulizia dei dataset. I dati riflettono il mondo così com'è, inclusi i suoi pregiudizi storici e le sue lacune. Senza un monitoraggio critico, i sistemi di AI agiscono come "black mirrors", amplificando le discriminazioni esistenti e creando echo chambers algoritmiche. Un rischio particolarmente insidioso è il cosiddetto model collapse: se i modelli futuri verranno addestrati su contenuti generati da AI attuali, si perderà progressivamente il contatto con la realtà empirica e diversificata, portando a una omologazione o, comunque, a una distorsione del pensiero e della conoscenza.

Una strategia per la data governance pone al centro la tracciabilità e la documentazione trasparente di ogni fase di pre-processing e di successiva validazione. Non basta avere dataset vasti; occorre che siano rappresentativi e che il loro utilizzo sia finalizzato al bene comune. L'investimento sui Data Spaces e sui Commons digitali non deve essere visto solo come un miglioramento dell'efficienza tecnica, ma come un atto di democrazia digitale. In ambiti sensibili come la salute o la giustizia, i dati devono essere gestiti con garanzie rafforzate, assicurando che la potenza computazionale dell'HPC non diventi uno strumento di esclusione. La tassonomia del rischio sviluppata in ICSC mutua da quella di IPCC per il rischio climatico e distingue tra hazard, esposizione e vulnerabilità. In questo modello, il rischio non è una proprietà della sola tecnologia. L'hazard è la possibilità che un modello di grandi dimensioni manifesti fallimenti tecnici, bias sistemici o allucinazioni. In un ambiente HPC, la potenza di calcolo agisce da moltiplicatore: un errore incorporato in un modello addestrato su scala massiva si propaga con una velocità e una magnitudo tali da rendere l'impatto potenzialmente catastrofico e difficile da invertire. L'esposizione riguarda invece quanto la società dipenda effettivamente dall'AI. Essa aumenta drasticamente quando settori vitali, come la pratica medica o la gestione delle reti energetiche, integrano i modelli nei propri workflow decisionali. Per mitigare questo fattore, ICSC sottolinea la necessità di mantenere meccanismi robusti di human-in-the-loop, assicurando che l'intelligenza umana resti il supervisore ultimo dei processi critici. La vulnerabilità, infine, rappresenta la predisposizione di un sistema a subire danni. Essa dipende dalla sensibilità dei gruppi coinvolti - ad esempio, quanto un algoritmo di riconoscimento facciale possa discriminare una minoranza - e dalla loro capacità adattiva. Un modello adeguato di rischio deve misurare quest'ultima attraverso il paradigma "Ready, Willing, Able": ovvero se i cittadini possiedono le competenze per interagire con l'AI (Ready), la fiducia per collaborare con essa (Willing) e l'effettivo accesso a infrastrutture e tutele etiche e giuridiche (Able).

In sintesi, il rischio tecnologico diventa un danno reale solo se colpisce soggetti esposti e privi di strumenti di difesa o mitigazione. Per questo, una strategia di innovazione responsabile deve agire su tre fronti: rendere i modelli tecnicamente più robusti per ridurre il pericolo, utilizzare ambienti di sperimentazione come le Regulatory Sandboxes per testare l'impatto sociale prima del rilascio, e investire in alfabetizzazione digitale per abbattere la vulnerabilità dell'ecosistema che utilizza i modelli.

Agentic AI e Augmented Algorithmic Agency: le nuove frontiere del rischio

La vera sfida di frontiera attuale è rappresentata dall'Agentic AI, ovvero sistemi che possiedono una

"Augmented Algorithmic Agency". Questi sistemi si distinguono sostanzialmente per tre proprietà chiave: la proattività nella scomposizione degli obiettivi (goal decomposition), la memoria persistente che permette di adattare le strategie nel tempo e l'interazione dinamica con strumenti esterni e altre AI agentici. Il rischio agentico emerge quando queste capacità portano a un distacco dalla supervisione umana continua. In un ambiente HPC, un agente intelligente che gestisce autonomamente workflow complessi può generare escalation non lineari, reagendo a dinamiche di mercato o geopolitiche in millisecondi, ben prima che un operatore umano possa intervenire.
Riferimenti regolatori come l'AI Act europeo, pur essendo pilastri normativi fondamentali, incontrano difficoltà nel "prendere le misure" a sistemi così evolutivi. Ad esempio, la conformità prevista dall'AI Act e la valutazione del rischio è spesso statica, basata su una fotografia del sistema al momento dell'immissione sul mercato. I sistemi agentici, tuttavia, deviano funzionalmente dal loro scopo originario attraverso il goal drift e l'apprendimento post-deployment. La conformità deve quindi diventare dinamica e deve sviluppare strumenti per modellare la transizione tra gli stati successivi dell'agente, misurando quanto il rispetto degli obblighi normativi (come la non-discriminazione o la sicurezza) rimanga stabile al variare dei dati e dei contesti operativi. Questo approccio responsabile mira a colmare il gap tra framework normativo e la pratica ingegneristica, fornendo ai regolatori europei modelli computazionali per controlli ex ante ed ex post realmente efficaci.

Sandboxes come infrastrutture HPC permanenti di governance
In questo contesto di evoluzione così veloce, le sandboxes regolatorie diventano uno strumento utile e una grande opportunità. Non devono essere intese come mere zone di "deregolamentazione", ma come ambienti controllati e protetti - fisici e digitali - dove sperimentare l'innovazione sotto la supervisione delle autorità. Le sandboxes devono trasformarsi in infrastrutture HPC permanenti, integrate con i propri Digital Twins settoriali. Immaginiamo, ad esempio, il gemello digitale di una città come Bologna: esso non deve servire solo per ottimizzare la mobilità o il consumo energetico, ma per fungere da sandbox sociale per valutare come l'introduzione di sistemi agentici influenzi l'accesso ai servizi pubblici o la mobilità dei cittadini più vulnerabili.
Le sandboxes per l'Agentic AI richiedono un cambio di prospettiva: dal test del "prodotto" al monitoraggio del "processo". Gli agenti dovranno essere testati iterativamente per individuare i "salti di rischio" inaspettati. Le infrastrutture HPC europee possono avere ruolo di leadership in questo campo, sfruttando la propria potenza per costruire ambienti di simulazione complessi che riflettano la realtà di interi settori industriali. In queste sandboxes, la supervisione umana non può essere solo un principio astratto, ma deve tradursi in interfacce tecniche certificate, come dashboard di audit che permettano di visualizzare non solo le azioni del sistema, ma l'intensità con cui esse toccano diritti e interessi specifici.

Tre esempi
AI e demografia: misurare la vulnerabilità per orientare le policy
L'analisi dei processi d'innovazione responsabile nello sviluppo delle tecnologie avanzate come l'AI evidenzia che l'impatto dell'AI sul mercato del lavoro e sulla società non è uniforme. In Italia, ad esempio, questo impatto si intreccia con sfide demografiche critiche: l'invecchiamento della popolazione, la denatalità e lo spopolamento delle aree rurali. Circa il 60% dei lavori nelle economie avanzate è esposto all'AI, ma questa esposizione può tradursi in complementarità o in sostituzione a seconda delle competenze e del territorio. Il framework "Ready, Willing, Able", anche in questo caso, può essere utilizzato per stimare la prontezza del Paese.
L'uso di simulazioni per mappare queste disuguaglianze è di grande utilità. L'obiettivo è supportare i decisori politici nella progettazione di interventi di formazione mirata e politiche di welfare che prevengano l'ampliamento dei divari territoriali. In parallelo, è auspicabile sviluppare "sandboxes

sociali" dove sperimentare l'uso di AI agentica, ad esempio, nel supporto domiciliare agli anziani o nei servizi educativi inclusivi. In questi contesti, l'AI non è un fattore di isolamento, ma uno strumento per rafforzare i legami comunitari e la dignità delle persone, garantendo che l'automazione ad alte prestazioni non lasci indietro nessuno.

Conoscenza e memoria: prevenire il debito cognitivo

Un'attenzione particolare deve essere rivolta agli effetti dell'AI generativa sul pensiero critico e sulla memoria umana. Il massiccio "offloading cognitivo" verso gli algoritmi comporta il rischio di una perdita di competenze di base e di una diminuzione della capacità di giudizio autonomo. Se deleghiamo costantemente all'AI la scrittura, l'argomentazione e la sintesi, rischiamo di contrarre un "debito cognitivo" che pagheremo in termini di omologazione e fragilità intellettuale.

È necessario definire standard per sistemi agentici "cognitivamente sani". Questo significa progettare la human-machine interaction affinché l'AI funga da partner di brainstorming e non da sostituto. I sistemi agentici devono essere progettati per stimolare domande, presentare contro-argomentazioni ed esplorare alternative, favorendo una "hybrid literacy" dove l'essere umano rimane il detentore ultimo dell'autonomia normativa. Un filone promettente è l'uso di agenti "etico-aware" che utilizzano ambienti di realtà virtuale per testare il giudizio morale degli utenti in contesti controllati, aiutando a estrarre e rappresentare quella conoscenza tacita valoriale che spesso va perduta nelle pipeline puramente numeriche.

Consumo energetico e sostenibilità

L'impronta ambientale dell'AI rappresenta una delle dimensioni più critiche del rischio sistemico di questa tecnologia avanzata. Se, come analizzato nelle sezioni precedenti, la potenza computazionale agisce da moltiplicatore per l'innovazione, essa funge anche da acceleratore per una domanda energetica senza precedenti. La letteratura scientifica più recente evidenzia come, ad esempio, il passaggio da GPT-3 a GPT-4 abbia comportato un incremento del consumo elettrico superiore a 40 volte, rendendo la "silicon scarcity" (scarsità di silicio) un problema secondario rispetto alla "electricity shortage" (carenza di energia).

In questo scenario, è necessario promuovere una transizione dal paradigma della "grandezza a ogni costo" verso lo sviluppo di modelli piccoli e specializzati (Small Language Models) e architetture tailored to the task, basate sulle seguenti argomentazioni strategiche:

Efficienza ed efficacia del "piccolo": L'addestramento di enormi modelli generali è intrinsecamente inefficiente. I dati dimostrano che l'aumento esponenziale dei parametri non garantisce incrementi proporzionali dell'accuratezza, incorrendo spesso in rendimenti marginali decrescenti. Al contrario, modelli più piccoli, addestrati su dataset verticali e di alta qualità, possono raggiungere prestazioni analoghe o superiori in domini specifici (come la sanità o l'energia) con una frazione del costo energetico. Strategie quali la quantizzazione, il pruning dei parametri e l'uso di architetture Sparse Mixture-of-Experts (MoE) permettono di attivare solo le porzioni di rete necessarie, riducendo drasticamente il consumo durante l'inferenza.

Sostenibilità dell'inferenza: Sebbene l'addestramento sia energeticamente intensivo, è la fase di inferenza (l'uso quotidiano del modello) a generare l'impatto maggiore nel lungo periodo, arrivando a coprire l'80-90% del consumo totale. L'adozione di modelli specializzati facilita il deployment in architetture cloud-to-edge: spostare l'elaborazione su dispositivi locali non solo protegge la privacy (sovranità del dato), ma riduce la pressione sui data center iperscalabili e sulle reti di distribuzione elettrica nazionali.

Valutabilità etica e regolatoria: I modelli generalisti di dimensioni colossali operano come "black box" quasi imperscrutabili, rendendo estremamente difficile la conformità dinamica richiesta dall'AI Act. Un modello "tailored", essendo più circoscritto, è intrinsecamente più auditable (verificabile). È più semplice mappare i bias, prevedere il goal drift e garantire la trasparenza dei dati di addestramento, riducendo il rischio di allucinazioni sistemiche che in un ambiente HPC si propagherebbero a

cascata.
Mitigazione dell'effetto rebound: È necessario riconoscere il paradosso per cui l'aumento dell'efficienza può portare a un aumento del consumo totale (effetto rebound). Pertanto, la strategia non deve limitarsi all'ottimizzazione del codice, ma deve includere una governance che privilegi l'uso di Carbon-Free Electricity (CFE) e la localizzazione dei carichi computazionali in periodi di alta produzione di energie rinnovabili, trasformando l'infrastruttura HPC in un consumatore energetico "consapevole".

Roadmap per un'innovazione responsabile
La costruzione di un'infrastruttura di innovazione responsabile non può esaurirsi nella semplice applicazione di norme tecniche, né in una verifica di conformità statica ed estemporanea. La sfida consiste nel trasformare il "rispetto delle regole" in un processo dinamico di apprendimento (regulatory learning). In questo senso, la roadmap per ICSC può poggiare su tre pilastri fondamentali che superano la visione tradizionale della compliance.
Dal controllo alla co-creazione: l'approccio multi-stakeholder
L'efficacia di una cornice etica non dipende solo dal rigore scientifico, ma dalla capacità di coinvolgere l'intero ecosistema attraverso processi di co-creazione. Non si tratta solo di far dialogare sviluppatori e regolatori, ma di strutturare un ambiente dove autorità competenti, fornitori di AI (con particolare attenzione a PMI e startup), infrastrutture, organismi di standardizzazione e società civile collaborino alla definizione degli standard. Questo approccio trasforma la regolamentazione da un "onere amministrativo" a un vantaggio competitivo: coinvolgendo le PMI e le startup sin dalle fasi embrionali dello sviluppo, si riducono le barriere all'ingresso nel mercato e si garantisce che le soluzioni tecnologiche siano "eticamente robuste" per costruzione, migliorando la certezza del diritto e l'attrattività degli investimenti.
Sperimentazione come infrastruttura di apprendimento
Le infrastrutture HPC, non devono essere considerate semplici zone di test, ma veri e propri laboratori di democrazia digitale. La sperimentazione pratica - supportata da sandbox regolatorie, da Digital Twin e da simulazioni ad alte prestazioni - consente un evidence-based regulatory learning. Per i regolatori: permette di comprendere l'evoluzione dei rischi (come il goal drift dei sistemi agentici) prima che diventino sistemici, adattando il quadro normativo in modo flessibile. Per gli sviluppatori: offre uno spazio protetto dove testare il bilanciamento tra principi conflittuali (es. privacy vs. efficienza) senza il rischio di sanzioni immediate, ma sotto una supervisione rigorosa che garantisce la tutela dei diritti fondamentali. Questo modello permette di superare la dicotomia tra "innovazione" e "regolazione", vedendo in quest'ultima un motore per l'innovazione di qualità.
Capacity Building e alfabetizzazione sistemica
Una roadmap responsabile deve infine investire sul capitale umano. Come evidenziato dalla necessità di nuovi modelli di formazione, è essenziale dotare sia il settore pubblico che quello privato di competenze interdisciplinari. Ridurre la vulnerabilità del sistema - il terzo fattore del paradigma Ready, Willing, Able - significa formare esperti capaci di condurre valutazioni d'impatto etico, sviluppatori consapevoli del "debito cognitivo" e dei costi energetici dei modelli, e decisori politici in grado di interpretare i dati delle simulazioni HPC. La leadership europea nell'AI non si misurerà solo sul numero di parametri dei modelli o sulla potenza di calcolo, ma sulla solidità di questo ecosistema di fiducia, dove la trasparenza e la tracciabilità sono garantite da infrastrutture di governance comuni e interoperabili.