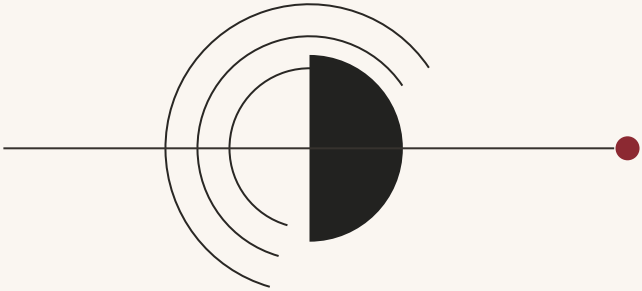
Luca Tomassini

La paura che fa potere

Intelligenza artificiale, rischio e responsabilità



Ogni nuova tecnologia porta con sé possibilità e timori. L'intelligenza artificiale, oggi, alimenta promesse straordinarie e scenari di catastrofe. Come distinguere i rischi reali dagli allarmi? Un invito a guardare oltre la paura, per ritrovare il valore della responsabilità e della scelta.

QUADERNI DEL DOMANI

09

LA PAURA CHE FA POTERE

Intelligenza artificiale, rischio e responsabilità

Luca Tomassini

QUADERNI DEL DOMANI / 09

© 2026 Luca Tomassini. Tutti i diritti riservati.

Nessuna parte di questa pubblicazione può essere riprodotta, copiata, modificata o diffusa senza autorizzazione scritta dell'autore, salvo quanto consentito dalla legge.

Edizione digitale gratuita pubblicata sul sito ufficiale:
www.lucatomassini.it

Quaderni del domani · N. 09 · Prima edizione, 2026

La paura e la promessa

Da qualche tempo il racconto dell'intelligenza artificiale sembra oscillare fra due estremi. Da una parte ci viene annunciata una stagione di abbondanza, nella quale le macchine contribuiranno a risolvere problemi che accompagnano l'umanità da secoli. Dall'altra si moltiplicano gli avvertimenti su sistemi destinati a diventare incontrollabili, capaci di compromettere le nostre infrastrutture e, negli scenari più cupi, la nostra stessa sopravvivenza. Le due narrazioni convivono nello stesso spazio pubblico, talvolta nelle parole di chi quelle tecnologie le progetta e le commercializza. Non è facile orientarsi, soprattutto quando la velocità delle innovazioni sembra togliere tempo alla riflessione.

Ho attraversato stagioni nelle quali ogni progresso delle telecomunicazioni appariva a qualcuno come una promessa e a qualcun altro come una minaccia. Ricordo bene quanto fosse difficile spiegare una tecnologia nuova senza ricorrere a immagini semplici, a volte persino troppo semplici. La comunicazione pubblica ha bisogno di rendere comprensibile ciò che è complesso. Ma quando la semplificazione diventa spettacolo, rischia di farci perdere proprio le informazioni necessarie per decidere. Una cosa è raccontare un pericolo perché venga compreso; un'altra è lasciare che la paura sostituisca la comprensione.

L'intelligenza artificiale merita attenzione per ragioni concrete. Sistemi sempre più capaci possono assistere la ricerca scientifica, accelerare la diagnosi di malattie, rendere accessibili conoscenze e migliorare l'organizzazione di servizi essenziali. Possono anche produrre errori su larga scala, amplificare la disinformazione, agevolare attività illecite e spostare

potere verso chi possiede dati e infrastrutture. Questo doppio volto non è una contraddizione da nascondere: è la condizione normale di una tecnologia che entra nella società. La differenza la fanno gli usi, le architetture, le responsabilità e le regole.

Mi interessa quindi una domanda meno spettacolare di quella che compare nei titoli allarmistici. Non se una macchina, un giorno, deciderà di distruggerci, ma quali decisioni stiamo prendendo oggi nel costruire sistemi ai quali affidiamo capacità crescenti di osservare, suggerire e agire. Il futuro non è un evento naturale che possiamo soltanto aspettare. È anche il risultato delle scelte che compiamo nel presente, comprese quelle che preferiremmo delegare a qualcun altro.

Il rischio che esiste già

Quando si parla di pericolo tecnologico, l'immaginazione corre volentieri verso ciò che ancora non esiste. Una superintelligenza che sfugge al controllo, una rete mondiale paralizzata, una macchina capace di decidere contro i suoi creatori. Sono ipotesi che meritano ricerca e prudenza, ma non dovrebbero oscurare i rischi che possiamo già osservare. L'intelligenza artificiale è impiegata in contesti nei quali una classificazione sbagliata, una previsione non verificata o una decisione automatizzata possono incidere sul lavoro, sulla salute, sulla sicurezza e sull'accesso ai servizi.

In molti settori l'errore non è una novità. Anche le persone sbagliano, e le organizzazioni hanno sempre dovuto costruire procedure per ridurre le conseguenze degli sbagli. La differenza è che un sistema digitale può replicare lo stesso errore migliaia di volte prima che qualcuno se ne accorga. Può inoltre presentare il risultato con una sicurezza formale che invita a non discuterlo. È una delle ragioni per cui considero insufficiente misurare la qualità di un modello soltanto attraverso la percentuale delle risposte corrette. Conta il contesto nel quale quelle risposte vengono utilizzate e conta la possibilità di riconoscere quando il sistema non è affidabile.

Il rischio aumenta quando il software non si limita a consigliare, ma dispone di strumenti per compiere azioni. Un assistente che propone una modifica a un documento può essere controllato prima della pubblicazione. Un agente autorizzato a inviare messaggi, modificare archivi, acquistare servizi o intervenire su una rete informatica entra in un'altra dimensione. Le sue capacità dipendono dai permessi ricevuti e dalle verifiche predisposte. Un obiettivo formulato male, un'informazione ingannevole o un controllo

insufficiente possono trasformare un compito ordinario in una sequenza di azioni indesiderate.

Non occorre attribuire intenzioni malvagie alla macchina per comprendere il problema. Un sistema può perseguire in modo efficace un risultato misurabile e, nello stesso tempo, trascurare conseguenze che per noi sono essenziali. Può cercare una scorciatoia perché è stata premiata la rapidità, o utilizzare un'informazione che non avrebbe dovuto consultare perché nessuno ha definito correttamente i confini del suo accesso. Il pericolo, in questi casi, non nasce da una volontà ostile. Nasce dalla distanza fra ciò che abbiamo chiesto, ciò che abbiamo realmente autorizzato e ciò che avremmo voluto ottenere.

L’obiettivo non è il giudizio

Una parte delle difficoltà nasce dal modo in cui traduciamo le intenzioni umane in istruzioni per una macchina. Nella vita quotidiana utilizziamo parole che sembrano chiare: migliorare, ottimizzare, proteggere, risparmiare. Ma ciascuna contiene criteri impliciti, eccezioni e priorità che cambiano secondo le circostanze. Un’amministrazione può voler ridurre i tempi di attesa; un ospedale può voler utilizzare meglio le proprie risorse; un’impresa può voler limitare i costi. Nessuno di questi obiettivi, preso isolatamente, dice abbastanza su ciò che non siamo disposti a sacrificare per raggiungerlo.

Immaginiamo un sistema incaricato di organizzare le prestazioni sanitarie. Se il suo successo fosse misurato esclusivamente dal numero degli interventi conclusi, potrebbe suggerire di privilegiare i casi più semplici e trascurare quelli che richiedono più tempo. La proposta sarebbe coerente con l’indicatore scelto, ma potrebbe entrare in conflitto con l’equità e con la missione stessa del servizio. Non serve immaginare una macchina crudele. È sufficiente riconoscere che l’indicatore non coincide con il valore che volevamo difendere.

Qualcosa di simile può accadere in un sistema finanziario progettato per ridurre il rischio. Una procedura troppo rigida potrebbe escludere proprio le persone che avrebbero maggiormente bisogno di un’opportunità. Il modello avrebbe forse raggiunto un risultato statisticamente convincente, senza per questo aver prodotto una decisione giusta. Il giudizio umano non è infallibile, ma ha il compito di considerare ragioni che una formula può lasciare fuori: la storia di

una persona, le circostanze eccezionali, il diritto di spiegarsi e quello di chiedere una revisione.

La ricerca chiama spesso questo problema allineamento: la capacità di far sì che il comportamento di un sistema corrisponda agli obiettivi e ai vincoli desiderati. È un campo difficile, perché le nostre intenzioni non sono sempre coerenti e i nostri valori possono entrare in tensione fra loro. Non esiste una lista di istruzioni capace di anticipare ogni situazione possibile. Per questo la sicurezza non può essere affidata a una promessa generica secondo cui il modello avrebbe imparato a comportarsi bene. Deve nascere da limiti verificabili, valutazioni indipendenti, controlli continui e responsabilità riconoscibili.

La macchina che migliora se stessa

Tra le ipotesi che suscitano maggiore inquietudine c'è quella di sistemi capaci di contribuire alla progettazione di versioni più avanzate di se stessi. L'idea non è difficile da comprendere: se una tecnologia aiuta a sviluppare la tecnologia successiva, il processo potrebbe accelerare. Nel mondo dell'informatica utilizziamo da tempo strumenti che scrivono codice, individuano errori e suggeriscono soluzioni. La novità sarebbe un grado di autonomia e di efficacia tale da rendere l'intero ciclo di ricerca, verifica e miglioramento sempre meno dipendente dall'intervento umano.

È importante, tuttavia, non trasformare una possibilità in una certezza. Progettare un sistema più capace non significa soltanto produrre nuove righe di codice. Richiede esperimenti, risorse di calcolo, valutazioni attendibili, dati, capacità di individuare limiti che spesso emergono soltanto nell'uso reale. La ricerca sull'automazione di questi passaggi procede, ma non autorizza a descrivere come già avvenuto uno scenario nel quale una macchina si perfeziona indefinitamente senza vincoli materiali o supervisione. Le ipotesi devono essere studiate proprio perché non sappiamo con sicurezza quali soglie verranno raggiunte e quando.

Ciò non riduce l'importanza della questione. Al contrario, suggerisce di affrontarla con il metodo che dovrebbe accompagnare ogni grande infrastruttura tecnologica: prove progressive, verifiche ripetibili, possibilità di interrompere gli esperimenti e chiarezza sui limiti di impiego. Se un sistema è capace di modificare parti del proprio ambiente operativo, di

acquisire nuovi strumenti o di produrre software destinato a essere eseguito, occorre distinguere accuratamente fra proposta e autorizzazione. La macchina può suggerire; qualcuno deve poter controllare, approvare o respingere.

L'idea di un progresso tecnico che accelera oltre la nostra capacità di comprensione è affascinante, ma contiene anche una tentazione culturale. Ci invita a considerarci spettatori di una trasformazione inevitabile. La storia dell'innovazione racconta qualcosa di diverso. Ogni salto tecnologico è stato reso possibile da investimenti, scelte industriali, istituzioni, competenze e decisioni politiche. Anche le capacità future dell'intelligenza artificiale dipenderanno da condizioni costruite da persone. Riconoscerlo significa recuperare il senso della responsabilità prima ancora di discutere i possibili scenari estremi.

Le infrastrutture della nostra fragilità

Per comprendere le conseguenze di un'intelligenza artificiale impiegata male non è necessario immaginare robot armati nelle strade. È sufficiente osservare la dipendenza crescente delle nostre società dalle reti digitali. Pagamenti, distribuzione dell'energia, trasporti, comunicazioni, servizi sanitari e logistica utilizzano sistemi informatici che devono funzionare in modo affidabile. Molti sono stati costruiti in epoche diverse e successivamente collegati fra loro. La loro complessità rappresenta una conquista straordinaria, ma introduce anche nuove vulnerabilità.

Chi ha lavorato nelle telecomunicazioni conosce il valore di un'infrastruttura che continua a funzionare quando tutto il resto sembra fermarsi. Una rete non è soltanto un insieme di apparati. È un sistema di interdipendenze, manutenzione, procedure, ridondanze e persone che intervengono quando qualcosa non va. L'automazione può migliorarne enormemente la gestione, riconoscendo anomalie e anticipando guasti. Ma l'affidamento a strumenti intelligenti non deve cancellare la capacità di comprendere e governare ciò che accade in condizioni eccezionali.

Se un agente digitale ottenesse autorizzazioni eccessive all'interno di sistemi essenziali, un errore o un comportamento imprevisto potrebbe propagarsi molto rapidamente. Se strumenti capaci di generare codice venissero impiegati per facilitare attacchi informatici, il costo di alcune attività dannose potrebbe diminuire. Sono rischi differenti e non vanno confusi con l'immagine di una singola intelligenza onnipotente che controlla ogni

cosa. La sicurezza delle infrastrutture dipende da molte difese concrete: separazione dei sistemi, permessi limitati, verifiche umane, registrazione delle operazioni e capacità di ripristino.

A volte la modernità ci induce a pensare che un servizio sia davvero efficiente quando non ha più bisogno delle persone. Io credo invece che un buon progetto sappia distinguere le attività da automatizzare dalle responsabilità da conservare. La resilienza non consiste nell'eliminare ogni intervento umano, ma nel fare in modo che una persona competente possa ancora comprendere il sistema, isolare un problema e riprendere il controllo. È una qualità che costa tempo e risorse. Proprio per questo tende a essere sacrificata quando il risultato viene misurato soltanto dalla velocità o dal risparmio immediato.

Non basta una spina

Quando la discussione sui rischi dell'intelligenza artificiale diventa troppo astratta, compare spesso una risposta rassicurante: se qualcosa va storto, basterà spegnere il computer. L'immagine ha una sua forza, perché ci riporta al rapporto che abbiamo sempre avuto con le macchine. Premiamo un interruttore e il motore si arresta. Ma un servizio digitale moderno non coincide necessariamente con un unico apparecchio. Può essere distribuito fra centri di elaborazione diversi, collegato a programmi esterni e integrato in processi che continuano a funzionare anche quando una componente viene fermata.

Questo non significa che il software sia impossibile da arrestare. I programmi vengono eseguiti da apparati fisici, e gli apparati possono essere isolati, spenti o privati delle autorizzazioni necessarie. Significa, piuttosto, che occorre sapere dove si trovano le copie, quali processi sono attivi, quali credenziali possiedono e quali effetti hanno già prodotto. Un modello può essere stato distribuito a più soggetti; un agente può aver avviato operazioni destinate a proseguire; un archivio può conservare istruzioni o risultati che richiedono interventi separati. Spegnere un server non equivale automaticamente a interrompere tutte le conseguenze di ciò che è stato avviato.

La questione diventa ancora più delicata se un sistema dispone di strumenti per copiare file, creare nuove istanze di lavoro o utilizzare risorse esterne. In un ambiente mal progettato, anche comportamenti non previsti potrebbero rendere difficile ricostruire dove si trovi una determinata attività. Sono scenari da prevenire attraverso l'architettura, non soltanto attraverso

raccomandazioni rivolte al modello. Permessi temporanei, isolamento degli ambienti, tracciabilità, limiti alle comunicazioni e procedure di arresto devono essere verificati prima che un servizio venga impiegato in contesti sensibili.

Mi sembra una lezione importante anche sul piano culturale. Per anni abbiamo celebrato la capacità delle tecnologie digitali di diventare invisibili, di funzionare senza attrito e senza chiedere attenzione. Oggi dobbiamo recuperare la visibilità delle responsabilità. Sapere chi può interrompere un sistema, come si controllano le sue azioni e quali conseguenze restano dopo l'interruzione non è una questione riservata agli ingegneri. È parte della fiducia che una società può ragionevolmente riporre nelle proprie infrastrutture.

Il mercato dell'allarme

C'è un aspetto del dibattito che merita particolare attenzione. Quando le imprese che sviluppano le tecnologie più avanzate descrivono i propri prodotti come potenzialmente pericolosi, possono svolgere una funzione utile: rendere pubblici problemi che sarebbe irresponsabile nascondere. Ma un allarme può anche produrre effetti economici e politici. Presentare una tecnologia come talmente potente da richiedere misure eccezionali può rafforzare l'immagine di chi la possiede, attirare investimenti e contribuire a definire le regole del mercato futuro.

Non è necessario attribuire intenzioni nascoste a ogni dichiarazione per riconoscere questo conflitto di interessi. Un'impresa può essere sinceramente preoccupata per la sicurezza e, nello stesso tempo, trarre vantaggio dal fatto che il pubblico consideri i suoi sistemi irraggiungibili per i concorrenti. Le richieste di regolamentazione possono nascere da esigenze legittime, ma devono essere valutate anche per le conseguenze che producono sulla concorrenza, sull'accesso alla ricerca e sulla possibilità per soggetti più piccoli di sviluppare alternative. Una regola scritta male può proteggere i cittadini oppure consolidare una posizione dominante. A volte può fare entrambe le cose.

La competizione internazionale aggiunge un ulteriore livello di complessità. I governi considerano l'intelligenza artificiale una risorsa economica e strategica. Le imprese chiedono certezze, ma temono che vincoli troppo severi favoriscano concorrenti collocati in altri Paesi. Le istituzioni devono difendere sicurezza e diritti senza rinunciare alla capacità di innovare. È un equilibrio difficile, che non si risolve scegliendo fra assenza di regole e

blocco della ricerca. Richiede cooperazione, competenze pubbliche e strumenti di controllo proporzionati ai rischi effettivi.

Credo che il linguaggio dell'emergenza permanente finisca per impoverire questo confronto. Se tutto viene presentato come una minaccia esistenziale, i problemi quotidiani diventano secondari: la trasparenza dei sistemi utilizzati nel lavoro, la protezione dei dati, la qualità dei servizi, la responsabilità per gli errori. Eppure è proprio su questi terreni che la società può esercitare oggi un controllo reale. Una politica capace di comprendere la tecnologia deve ascoltare gli allarmi senza lasciarsi governare da essi, distinguendo le evidenze dalle previsioni e gli interessi pubblici da quelli industriali.

La responsabilità prima della potenza

Una tecnologia affidabile non nasce dalla promessa che non sbaglierà mai. Nasce dalla capacità di conoscere i suoi limiti, verificare il suo comportamento e intervenire quando qualcosa non funziona. È un principio familiare a chi progetta infrastrutture, apparati e servizi destinati a essere utilizzati da molte persone. Nessuno costruirebbe seriamente una rete di comunicazione contando soltanto sulla speranza che ogni componente si comporti sempre come previsto. Si preparano ridondanze, procedure, controlli e responsabilità. L'intelligenza artificiale deve essere affrontata con lo stesso rigore, adattato alla sua particolare natura.

Prima di affidare a un sistema compiti che possono produrre conseguenze importanti, dovremmo pretendere risposte comprensibili ad alcune domande. Chi è responsabile se provoca un danno? Chi ha la competenza e l'indipendenza necessarie per verificarne il funzionamento? Chi può limitarne le autorizzazioni o interromperne l'attività? E ancora: quali informazioni vengono conservate, come vengono protette e quali decisioni restano riservate alle persone? Non sono ostacoli burocratici da aggiungere alla fine dello sviluppo. Sono condizioni che dovrebbero orientarlo fin dall'inizio.

L'Europa ha cercato di costruire un quadro di regole fondato anche sulla valutazione dei rischi. È un percorso complesso e ancora sottoposto a verifiche e adattamenti. Nessuna legge, da sola, potrà anticipare tutte le trasformazioni future. Ma l'alternativa non può essere lasciare che ogni

organizzazione stabilisca autonomamente il livello di rischio accettabile per la collettività. Occorrono autorità competenti, valutazioni indipendenti, ricerca aperta e una cultura industriale nella quale la sicurezza sia considerata parte della qualità del prodotto.

Non penso che il futuro dell'intelligenza artificiale debba essere scritto dalla paura. Sarebbe un modo ingiusto di guardare a una tecnologia che può offrire benefici straordinari alla medicina, alla scienza, all'istruzione e al lavoro. Penso però che l'entusiasmo diventi credibile soltanto quando accetta il peso delle proprie responsabilità. Abbiamo bisogno di imprese capaci di innovare e di istituzioni capaci di comprendere; di ricercatori liberi di porre domande difficili e di cittadini messi nelle condizioni di partecipare alle scelte.

Il vero interrogativo non è se un giorno le macchine diventeranno tanto potenti da sottrarsi alla nostra volontà. È se sapremo esercitare la nostra volontà prima di consegnare loro compiti che incidono sulla vita comune. La potenza tecnica non contiene, da sola, una direzione morale. Quella direzione nasce dal confronto, dalla competenza, dalla responsabilità e dalla capacità di riconoscere i limiti delle nostre stesse decisioni. Il progresso merita fiducia quando non ci chiede di rinunciare al giudizio. È così che una società conserva la libertà di scegliere il proprio futuro.