Centres for European Policy Network
FREIBURG | BERLIN | PARIS | ROMA

cepAdhoc

N. 15 | 2022 13 dicembre 2022

L’intelligenza artificiale come rischio sistemico in tempi di “policrisi”

Il pericolo della previsione algoritmica in contesti sconosciuti

Anselm Küsters

La tanto decantata potenza dell'intelligenza artificiale (AI) dipende, in molti campi di applicazione, dai dati sulla base dei quali essa viene sviluppata. Dal momento che il nostro mondo, sempre più dinamico e interdipendente, si discosta dalle osservazioni del passato utilizzate per la progettazione e lo sviluppo della AI, va da se che i sistemi di AI rischieranno di diventare meno affidabili. Per aumentare la resilienza dell'Unione europea (UE), la proposta di normativa europea sull'AI dovrebbe quindi essere adattata per tenere conto della maggiore fallibilità dei sistemi di AI in tempi di “policrisi”, soprattutto quando è il settore pubblico ad utilizzarli.

- In tempi di disordine globale, alcuni esprimono la speranza che i sistemi di AI predittivi possano aiutare, ad esempio, a prevenire la prossima pandemia o a risparmiare risorse per combattere il cambiamento climatico.
- Tuttavia, poiché gli eventi del mondo reale si discostano sempre più dai dati utilizzati per fare previsioni algoritmiche, i sistemi di AI diventeranno meno affidabili e richiederanno una crescente supervisione. Diversi esempi recenti illustrano questa tendenza preoccupante.
- Piuttosto che mettere a punto gli algoritmi per le crisi attuali, i responsabili politici dovrebbero stabilire condizioni quadro solide che, pur non massimizzando l'efficienza, consentano alle economie europee guidate dalla tecnologia di operare in modo ragionevole quando la prossima crisi colpirà. Per questo motivo, il Center for European Policy Network (CEP) chiede che la proposta di legge sull'AI preveda una classificazione dei sistemi di AI in base al contesto e che la divulgazione dei risultati avvenga tramite *audit* mirati.

Indice

1 Introduzione: AI in soccorso?

Il potere dell'intelligenza artificiale (AI) è indubbiamente affascinante, quindi non sorprende che, in tempi di disordine globale, alcuni esprimano la speranza che gli algoritmi possano aiutare a prevenire la prossima pandemia,[1] a risparmiare risorse per combattere il cambiamento climatico[2] o persino a realizzare una società più equa ed inclusiva.[3]Ciò che spesso si dimentica è che qualsiasi sistema di AI è fondamentalmente limitato dal tipo di dati sulla base dei quali viene predisposto. Come regola generale, i moderni algoritmi di apprendimento automatico funzionano meglio se sviluppati su grandi quantità di dati di alta qualità e funzionano peggio quando i dati sottostanti sono imprecisi, incompleti, irrilevanti, non validi, non aggiornati o incoerenti.[4]Che cosa implica questa relazione fondamentale per i tempi correnti di disordine globale, caratterizzati da molteplici crisi che vanno dalla guerra della Russia contro l'Ucraina, alla scarsità di energia, all'impennata dell'inflazione e al caos climatico?

Poiché gli eventi del mondo reale si discostano sempre più dai dati utilizzati per formulare previsioni algoritmiche, i sistemi di AI diventeranno meno affidabili e richiederanno una crescente supervisione. È fondamentale riconoscere i limiti di questi sistemi, che derivano da confini epistemici. Pertanto, sia i politici che le autorità di regolamentazione devono affrontare la triste realtà che questa tecnologia di per sé non ci salverà da recessioni, guerre o cambiamenti climatici. È necessario, invece, un quadro normativo resiliente, a partire dalla legge sull'intelligenza artificiale attualmente prevista, e una buona dose di buon senso.

2 Background: il cigno nero incontra il rinoceronte grigio

Rendersi conto della limitata praticabilità dei modelli quantificati in tempi di cambiamenti improvvisi è un punto di partenza prezioso per riflettere sul ruolo dei sistemi di intelligenza artificiale nel mondo caotico di oggi. L'apparente limitazione dei modelli economici tradizionali durante i primi mesi della crisi finanziaria del 2007/08 è un caso emblematico: "Di fronte alla crisi", lamentava Jean-Claude Trichet, allora presidente della Banca Centrale Europea, nel novembre 2010, "ci siamo sentiti abbandonati dagli strumenti convenzionali". Questi strumenti erano stati formulati e messi a punto durante un periodo di volatilità macroeconomica straordinariamente bassa, noto come "Grande moderazione"[5], e sono stati poi sovraccaricati dalle turbolenze di mercato che si stavano diffondendo. Senza una chiara guida da parte dei quadri analitici esistenti, le banche centrali di tutto il mondo hanno guardato alla storia ciclica delle crisi per trovare lezioni qualitative che potessero essere apprese.[6]

---

1 Makin, S. (2022), Could an algorithm predict the next pandemic?, Nature Outlook (26.10.2022), https://www.nature.com/articles/d41586-022-03358-4.

2 Miller, K. (2022), Building Intelligent Agents to Reach Net Zero 2050, HAI Stanford University (3.10.2022), https://hai.stanford.edu/news/building-intelligent-agents-reach-net-zero-2050.

3 Lobel, O. (2022), The Equality Machine: Harnessing Digital Technology for a Brighter, More Inclusive Future, New York: Public Affairs.

4 Pogrebivsky, S. (2021), The 6 Attributes of High-Quality Data, DataGroomr (8.7.2021), https://datagroomr.com/understand-the-6-attributes-of-high-quality-data/.

5 Bean, C. (2010), The Great Moderation, the Great Panic, and the Great Contraction, Journal of the European Economic Association8 (2-3), pp. 289-325.

6 Eichengreen, B. (2015), Hall of mirrors: the Great Depression, the Great Recession, and the uses—and misuses—of history, Oxford: Oxford University Press.

L'osservazione che la maggior parte di questi economisti ha finito per utilizzare analogie con la "Grande depressione" degli anni trenta,[7] illustra che i punti di riferimento utili per l'orientamento sono pochi e lontani quando si cercano lezioni dal passato in tempi di crisi. Le statistiche della storia economica dimostrano che le grandi crisi finanziarie sono eventi rari: In media, le crisi si verificano ogni 25 anni e le nuove recessioni iniziano in genere ogni otto anni.[8] Questo limita fortemente qualsiasi sforzo empirico, compreso l'"addestramento" di sistemi di intelligenza artificiale, per esempio, per prevedere tali eventi. Inoltre, anche se esistono dati su tali crisi, essi sono quasi per definizione "sporchi", cioè mancano di alcune informazioni o non misurano la realtà con la stessa precisione di un contesto normale. I modelli progettati o "addestrati" tenendo conto dei "tempi normali" potrebbero quindi iniziare a fallire quando si verificano forti shock esterni che segnalano l'inizio di "tempi anormali".

A causa della natura intrinseca del rischio, questo problema non può essere interamente previsto o aggirato, un fenomeno che ha portato l'ex trader di Wall Street Nassim Nicholas Taleb a coniare il termine "cigni neri".[9] I cigni neri sono eventi imprevisti e imprevedibili con conseguenze estreme e, come tali, non possono mai essere coperti completamente dai nostri modelli. Oltre alla crisi finanziaria del 2007/08, anche eventi più recenti, come la pandemia di Covid e l'attacco della Russia all'Ucraina, rientrano in questa categoria. L'utilizzo di sistemi di AI per supportare i decisori in tali condizioni anomale comporta tipicamente "criteri contrastanti e incomparabili", ad esempio il costo rispetto al benessere umano.[10] Inoltre, gli economisti hanno usato l'espressione "rinoceronti grigi" per indicare rischi ben noti e in lento movimento che possono amplificare gli shock esterni, come l'attuale alto livello di indebitamento delle famiglie o il cambiamento climatico globale. [11] Mentre i cigni neri e i rinoceronti grigi complicano qualsiasi previsione, compreso il ragionamento umano, le applicazioni di AI basate sulla matematica che circolano al giorno d'oggi sono molto più opache dei sistemi di previsione tradizionali, tipicamente al di là di ogni contestazione o appello, e possono amplificare i cicli di feedback negativi su grandi distanze in brevi periodi.[12] L'aumento dell'uso di queste applicazioni crea quindi nuovi rischi sistemici.[13]

Le cose si complicano ulteriormente quando diversi shock colpiscono contemporaneamente e stabiliscono interconnessioni che non sono generalmente previste. Lo storico dell'economia Adam Tooze, ha recentemente descritto lo stato attuale del mondo come una "policrisi", denotando un'interazione di shock disparati che, nel suo insieme, è peggiore della somma delle sue parti. [14] Il quadro che emerge da questa breve indagine è chiaro: la presenza di molteplici cigni neri e rinoceronti grigi nell'attuale policrisi ha molto probabilmente un impatto sull'utilità dei nostri algoritmi che sono, per forza di cose, "addestrati" su dati passati, che potrebbero ora diventare rapidamente obsoleti. Il problema dell'"addestramento" dei modelli su insiemi di dati più ristretti

7 Küsters, A. (2022). Applying Lessons from the Past? Exploring Historical Analogies in ECB Speeches through Text Mining, 1997-2019. International Journal of Central Banking 18 (1), pp. 277-329.

8 Paul, P. (2019), Modeling Financial Crises, FRBSF Economic Letter (2019-08).

9 Taleb, N. (2007), The Black Swan: The Impact of the Highly Improbable, New York: Random House.

10 Mostaghim, S. (2020), AI to the Rescue: Life-and-Death Decision-Making under Conflicting Criteria, Project website (undated), https://forschung-sachsen-anhalt.de/project/ai-rescue-life-death-decision-making-23474.

11 MarjaNykänen, M. (2022), Black swans and grey rhinos – lessons of crises on macroprudential policy, Opening remarksat the Conference on Systemic Risk Analytics, Helsinki (5.5.2022), https://www.bis.org/review/r220509c.htm.

12 O'Neil, C. (2016), Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy, New York: Crown Publishers.

13 Galaz, V. et al. (2021), Artificial intelligence, systemic risks, and sustainability, Technology in Society 67, 101741.

14 Tooze, A. (2022), Welcome to the world of the polycrisis, FT (28.10.2022), https://www.ft.com/content/498398e7-11b1-494b-9cd3-6d669dc3de33.

rispetto alla popolazione che in ultima analisi sono destinati a riflettere è una forma di cosiddetta "fuga di dati", che minaccia l'affidabilità dell'apprendimento automatico in tutte le discipline.[15] Ancor più grave è il fatto che l'utilizzo di sistemi di AI predittivi per eliminare le incertezze in tempi di crisi, potrebbe addirittura finire di generarne di più.

3 Esempi: perché l'AI può non funzionare in tempi "anormali"

I pericoli possono essere dimostrati riportando brevemente l'orologio ai primi mesi della pandemia di Covid. Non appena sono iniziati i primi blocchi, gli economisti si sono affannati a cercare di capire lo stato attuale dell'economia globale affidandosi a nuovi dati in tempo reale, poiché le misure standard erano troppo lente o inaffidabili. Invece di aspettare le stime ufficiali sull'inflazione e sulla disoccupazione, hanno attinto sempre più spesso a indicatori precedentemente oscuri, come le statistiche sulla mobilità di Apple o Google, o i dati sulle prenotazioni dei ristoranti, per farsi un'idea dell'attività economica complessiva.[16] Mentre queste misure sono riuscite a catturare il cambiamento del comportamento dei consumatori durante la pandemia, i sistemi di intelligenza artificiale addestrati sui dati che si adattano ai vecchi modelli di spesa hanno invece immediatamente incontrato dei problemi.

Prendiamo il caso di Fair Isaac Corp (FICO), uno sviluppatore di software con sede negli Stati Uniti, i cui importanti strumenti di intelligenza artificiale per il rilevamento delle frodi sulle carte di credito e di debito sono utilizzati dalle grandi banche per prendere decisioni sui prestiti. All'inizio della pandemia, sulla base dell'esperienza passata, questi strumenti si aspettavano molti più acquisti di persona che virtuali, una situazione che alla fine ha portato a segnalare come problematiche un gran numero di transazioni virtuali. Ciò significava che l'algoritmo sottostante consigliava di negare milioni di acquisti legittimi, mentre i consumatori in isolamento cercavano ardentemente di assicurarsi beni di prima necessità online.[17] Nello stesso periodo, in Cina, gli strumenti di valutazione del credito automatizzati dall'intelligenza artificiale di grandi operatori come Ant Group, che avevano costantemente mantenuto i tassi di insolvenza all'uno per cento, hanno funzionato molto meno agevolmente.[18] Ciò è dovuto in parte all'aumento dello stress finanziario durante la prima pandemia. Tuttavia, i nuovi problemi probabilmente riflettevano anche il fatto che i punteggi di Ant non si basavano solo su fattori tradizionali come la storia creditizia, ma utilizzavano anche criteri molto più ampi, come gli hobby e le preferenze di acquisto di un utente - in altre parole, misure comportamentali che hanno subito cambiamenti drammatici a causa dell'autoisolamento.

In altri contesti di crisi, come la salute o il cambiamento climatico, che tendono ad assomigliare più a rischi di "rinoceronte grigio" che di "cigno nero", il problema non è il cambiamento improvviso degli indicatori critici, ma la mancanza di dati di alta qualità che indebolisce la previsione algoritmica. A causa di problemi relativi alla dimensione e alla composizione dei dati utilizzati per "addestrare" i modelli di IA per l'assistenza sanitaria, questi ultimi non sono così accurati nel prevedere le malattie

[15] Gibney, E. (2022), Could machine learning fuel a reproducibility crisis in science?, Nature Outlook (26.7.2022), https://www.nature.com/articles/d41586-022-02035-w.

[16] The Economist (2020), Why real-time economic data need to be treated with caution (23.7.2020), https://www.economist.com/finance-and-economics/2020/07/23/why-real-time-economic-data-need-to-be-treated-with-caution.

[17] Dave, P. (2022), When the AI goes haywire, bring on the humans, Reuters (13.10.2022), https://www.reuters.com/technology/when-ai-goes-haywire-bring-humans-2022-10-13/.

[18] Chorzempa, M. (2022), The Cashless Revolution: China's Reinvention of Money and the End of America's Domination of Finance and Technology, New York: Public Affairs, pp. 98f., 103.

come suggeriscono i rapporti.[19] Allo stesso modo, poiché i Paesi variano nella qualità e nella quantità di dati meteorologici e altri dati geografici raccolti, i modelli di previsione del rischio climatico per alcune aree potrebbero essere errati.[20] In genere, le comunità piccole e a basso reddito che affrontano le maggiori perdite finanziarie e i maggiori rischi in un mondo di cambiamenti climatici non dispongono del tipo di dati climatici di alta qualità necessari per formare sistemi di allerta precoce con la capacità di rispondere rapidamente.[21] Ad esempio, i ricercatori del settore energetico criticano il fatto che gli attuali piani di lotta al cambiamento climatico basati sui dati falliscono a causa della mancanza di dati e modelli specifici per l'Africa.[22] Questa divisione tra chi ha i dati e chi non li ha ostacola l'azione collaborativa [23]e può rafforzare le disuguaglianze tra le comunità, poiché un sistema di intelligenza artificiale adatterà le sue raccomandazioni alle aree in cui può attingere a una maggiore quantità di dati.

In effetti, gli esperti si rendono sempre più conto che l'utilizzo di analisi predittive in situazioni di complessità, disordine o dati insufficienti, potrebbe addirittura rafforzare le tendenze negative sottostanti. Ad esempio, per gestire le conseguenze della crisi degli oppioidi, gli Stati Uniti si affidano a un importante algoritmo di rischio di tossicodipendenza che, come hanno ora documentato i ricercatori, sembra peggiorare la situazione.[24] Gli strumenti di screening dei pazienti basati sull'intelligenza artificiale sono intrinsecamente difettosi, in quanto non sono in grado di gestire costrutti complessi come la salute umana e, pertanto, spesso negano il trattamento ai pazienti più vulnerabili o che rappresentano casi clinicamente complessi. La consapevolezza del potenziale di questi sistemi automatizzati di trasformarsi in "armi di distruzione matematica"[25] ha conseguenze di vasta portata quando si pensa alla regolamentazione: "Se le analisi predittive stanno in parte creando la realtà che pretendono di prevedere", come ha recentemente concluso Carissa Véliz, ricercatrice dell'Università di Oxford, "allora sono in parte responsabili delle tendenze negative che stiamo vivendo nell'era digitale, dall'aumento delle disuguaglianze alla polarizzazione, alla disinformazione e ai danni a bambini e adolescenti".[26] L'attuale crisi multipla esacerberà questi problemi fondamentali alla base delle previsioni algoritmiche.

Un ultimo esempio dal settore della sicurezza, illustra questa preoccupante intuizione. Un recente studio pubblicato dall'Agenzia dell'Unione europea per i diritti fondamentali critica l'aumento dell'uso di sistemi di AI per la polizia predittiva.[27] Anche in questo caso, il problema degli strumenti basati sull'AI - in questo caso per le forze dell'ordine - è che di solito si basano su dati storici e quindi

[19] Berisha, V. / Julie Liss, J. (2022), AI in Medicine Is Overhyped, Scientific American (19.10.2022), https://www.scientificamerican.com/article/ai-in-medicine-is-overhyped/.

[20] Center for Data Innovation (2022), How Does the Data Divide Impact Global Policy Challenges?, Online Panel Discussion (7.12.2022), https://datainnovation.org/2022/12/how-does-the-data-divide-impact-global-policy-challenges/.

[21] Naddaf, M. (2022), Climate change is costing trillions – and low-income countries are paying the price, Nature News (7.11.2022), https://www.nature.com/articles/d41586-022-03573-z.

[22] Mutiso, R. (2022), Net-zero plans exclude Africa, Nature World View (2.11.2022), https://www.nature.com/articles/d41586-022-03475-0.

[23] Diebold, G. (2022), Data Divide or Digital Divide? Or Both?, ISE Magazine (7.11.2022), https://www.isemag.com/industry-trends-and-research/article/14284950/data-divide-or-digital-divide-or-both.

[24] Szalavitz, M. (2021), The Pain Was Unbearable. So Why Did Doctors Turn Her Away?, Wired (11.8.2021), https://www.wired.com/story/opioid-drug-addiction-algorithm-chronic-pain/.

[25] O'Neil, C. (2016), Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy, New York: Crown Publishers.

[26] Véliz, C. (2021), If AI Is Predicting Your Future, Are You Still Free?, Wired (27.12.2021), https://www.wired.com/story/algorithmic-prophecies-undermine-free-will/.

[27] FRA (2022), Bias in algorithms – Artificial intelligence and discrimination, Luxembourg: Publications Office of the EuropeanUnion, https://fra.europa.eu/en/publication/2022/bias-algorithm.

potenzialmente obsoleti e distorti, mentre l'AI apprende i modelli emergenti. I "cicli di feedback" che ne derivano sono particolarmente dannosi in caso dicrisi multiple. Ad esempio, una minore presenza di polizia in un'area si tradurrà di solito in un minor numero di crimini, che porteranno a un'area ancora meno sorvegliata in futuro, poiché l'AI raccomanda di allocare le risorse ad altri siti che ritiene più critici. Questo, a sua volta, rende l'area specifica più vulnerabile agli shock esterni, come gli attacchi alle infrastrutture critiche o al terrorismo, poiché il modello di AI incorpora solo i dati sulle accuse criminali passate e non può pensare strategicamente ad altri aspetti rilevanti, come la posizione di una stazione ferroviaria frequentemente visitata o di un fornitore di energia critico. Lo stesso problema può essere rilevato per i sistemi di AI predittivi utilizzati per la migrazione, l'asilo e il controllo delle frontiere, che diventano sempre più importanti in tempi di mobilità umana indotta dai cambiamenti climatici, ma le cui valutazioni errate "hanno conseguenze significative per la preparazione degli Stati membri, ma anche per la probabilità che gli individui possano accedere alla protezione internazionale". [28] In altre parole, l'utilità degli strumenti di previsione basati sull'AI è limitata a domini ristretti - ad esempio, la finanza, l'ambiente, la medicina o la sicurezza - escludendo tutti gli altri. Come ha osservato Martin Wolf, pensare al mondo in “silos intellettuali” può essere efficiente in un mondo ragionevolmente stabile, ma fallirà inevitabilmente in una situazione di crisi multiple.[29]

4 Outlook: come mitigare I rischi di una falsa predizione degli algoritmi

In una certa misura, la raccolta di nuovi dati, l'aumento della condivisione dei dati a livello globale e l'imposizione di standard di dati comuni possono migliorare la copertura geografica e cronologica, portando così a modelli quantificativi migliori. Un'altra possibile strada potrebbe essere l'apprendimento per rinforzo, che non dipende da serie di dati esterni ma da informazioni create durante lo sviluppo del modello. Alcuni ricercatori sperano di sviluppare sistemi di intelligenza artificiale più resistenti, basati su un tipo di ragionamento statistico noto come "pianificazione sequenziale sotto incertezza".[30] Tuttavia, queste alternative non sono una panacea, poiché tutti gli insiemi di dati sono in qualche modo imperfetti e i "cigni neri" potrebbero verificarsi in qualsiasi momento. Inoltre, la crescente velocità e la crescente interconnessione delle crisi rendono praticamente impossibile mantenere i dati aggiornati con una frequenza sufficiente. È interessante notare che anche gli esperti che sperano che l'apprendimento automatico possa aiutare a identificare i virus più pericolosi ammettono che gli strumenti predittivi non possono prevenire la prossima pandemia.[31] Allo stesso modo, gli esperti suggeriscono che non esistono "pallottole d'argento" per produrre modelli di intelligenza artificiale affidabili e clinici.[32]

---

[28] Access Now et al. (2022), Open Letter: The EU AI Act must protect people on the move, https://edri.org/wp-content/uploads/2022/12/Open-letter_EU-AI-Act_migration_December-2022.pdf, p. 3.

[29] Wolf, M. (2022), How to think about policy in a polycrisis, FT (29.11.2022), https://www.ft.com/content/a1918fec-2c8f-4051-ad78-c300b0fc9adb.

[30] Miller, K. (2022), Building Intelligent Agents to Reach Net Zero 2050, HAI Stanford University (3.10.2022), https://hai.stanford.edu/news/building-intelligent-agents-reach-net-zero-2050.

[31] Makin, S. (2022), Could an algorithm predict the next pandemic?, Nature Outlook (26.10.2022), https://www.nature.com/articles/d41586-022-03358-4.

[32] Berisha, V. / Julie Liss, J. (2022), AI in Medicine Is Overhyped, Scientific American (19.10.2022), https://www.scientificamerican.com/article/ai-in-medicine-is-overhyped/.

In definitiva, la risposta alla domanda se i nostri modelli possano mai sperare di funzionare in un mondo sempre più complesso si riduce a stabilire se stiamo solo insegnando agli algoritmi a ripetere cose che già conoscono o se sono in grado di apprendere principi completamente nuovi - una questione che è ancora molto dibattuta tra gli scienziati informatici e i filosofi e che probabilmente non sarà risolta a breve. Data l'insicurezza e l'ambivalenza, non bisogna affidare ai tanto sbandierati algoritmi la soluzione delle crisi attuali o la capacità di resilienza. Bisogna invece adottare un approccio più sfumato. La migliore difesa contro i rischi della previsione algoritmica in tempi di disordine globale consiste nel mantenere un panorama normativo semplice e solido. Le regole non devono essere così dettagliate e onnicomprensive da impedire la nascita di start-up innovative. Al contrario, esse dovrebbero definire le condizioni quadro in modo che il mercato possa separare gli strumenti di previsione basati sull'IA sensati da quelli non sensati nel lungo periodo, senza che queste applicazioni possano causare danni alla società nel breve periodo.

Questo solido quadro normativo dovrebbe prevedere l'obbligo per un maggior numero di team aziendali di verificare le prestazioni dell'AI per evitare problemi come quelli segnalati da FICO. Anche se la nuova legge europea sull'AI richiederà un certo monitoraggio, c'è ancora molto da fare: un'indagine McKinsey condotta nel 2021 su 1.843 aziende suggerisce che la maggior parte di esse non controlla regolarmente i programmi basati sull'AI dopo averli lanciati.[33]Anche le aziende che dispongono di team responsabili per l'AI non investono abbastanza, con il risultato che le persone che lavorano nel settore ne risentono. [34] Le aziende dovrebbero essere obbligate per legge a pubblicare i loro audit sull'AI, il che incentiverebbe ulteriormente l'impiego di maggiori risorse per indagare sui danni dell'AI. Sebbene questi audit siano stati suggeriti per aumentare la responsabilità sull'uso degli algoritmi, la loro corretta attuazione richiede una guida normativa su standard sufficienti e pratiche comuni. Poiché la legge sull'AI e la legge sulla responsabilità dell'AI attualmente previste dall'UE imporranno alle aziende di documentare come stanno attenuando i danni, sono necessari standard chiaramente definiti e maggiori risorse umane e tecniche per evitare un ulteriore "esaurimento" nel settore dell'audit, che potenzialmente aggraverebbe, anziché attenuare, i danni dell'AI.[35] Una possibile soluzione potrebbe essere quella di imporre la revisione delle componenti critiche dei risultati da parte di esperti.[36]

Un punto di partenza cruciale per l'adeguamento sarà rappresentato dai negoziati in formato "trilogo" del prossimo anno tra Commissione, Consiglio e Parlamento europeo sull'"AI Act", l'iniziativa faro europea per la regolamentazione dei sistemi di apprendimento automatico nell'era digitale. Le sue regole dovrebbero essere adattate per tenere conto della maggiore fallibilità dei sistemi di AI in tempi di "policrisi". In generale, l'imprevedibilità degli eventi "cigno nero" e "rinoceronte grigio" suggerisce che un approccio puramente basato sul rischio, potrebbe non essere sufficiente, poiché non possiamo conoscere il rischio complessivo di un determinato sistema.[37]

---

[33] McKinsey Analytics (2021), The state of AI in 2021 (December).

[34] Heikkilä, M. (2022), Responsible AI has a burnout problem, MIT Technology Review (28.10.2022), https://www.technologyreview.com/2022/10/28/1062332/responsible-ai-has-a-burnout-problem/.

[35] Costanza-Chock, S. /Raji, I. / Buolamwini, J. (2022), Who Audits the Auditors? Recommendations from a Field Scan of the Algorithmic Auditing Ecosystem, ACM Conference on Fairness, Accountability, and Transparency, pp. 1571–1583.

[36] Costanza-Chock, S. /Raji, I. / Buolamwini, J. (2022), Who Audits the Auditors? Recommendations from a Field Scan of the Algorithmic Auditing Ecosystem, ACM Conference on Fairness, Accountability, and Transparency, p. 1579.

[37] The version from 25 November 2022 can be found here: https://data.consilium.europa.eu/doc/document/ST-14954-2022-INIT/de/pdf.Siehe auch: Kullas, M. / Harta, L. (2021), Europäisches Gesetz über Künstliche Intelligenz Kurzfassung, cepAnalyse zu COM2021 206(13.12.2021), https://www.cep.eu/eu-themen/details/cep/europaeisches-gesetz-ueber-kuenstliche-intelligenz-cepanalyse-zu-com2021-206.html.

Tuttavia, se si accetta l'approccio basato sul rischio dell'attuale bozza, come un dato di fatto, si potrebbero allora incorporare i pericoli che sorgono in tempi di crisi multiple classificando una percentuale maggiore di sistemi guidati dall'AI come "ad alto rischio" ogni volta che l'attuale clima economico o politico suggerisce che i loro dati di "addestramento" potrebbero rivelarsi fuori dalla realtà. I sistemi ad alto rischio, ai sensi dell'AI Act, sono quelli che possono avere un impatto significativo sulle possibilità di vita di un utente e sono quindi tenuti a certificare, tra l'altro, dati di formazione di alta qualità, un'adeguata supervisione umana e test di accuratezza e robustezza. Il più recente testo di compromesso per l'AI Act elenca otto tipi concreti di sistemi che rientrano in questa categoria, come i sistemi automatizzati per la formazione professionale o le forze dell'ordine.[38]Tuttavia, uno schema di classificazione sensibile al contesto consentirebbe di includere un maggior numero di sistemi di AI in questo regime normativo in tempi di crisi multiple, garantendo così che questi sistemi debbano soddisfare standard più elevati in termini di solidità dei dati.

Ad esempio, si potrebbe imporre che tutte le applicazioni di AI che rientrano nella categoria dei "sistemi a rischio limitato", che in tempi normali richiedono solo obblighi di trasparenza rudimentali, debbano soddisfare gli obblighi aggiuntivi per i sistemi "ad alto rischio" quando gli shock esterni in corso, come guerre, pandemie o perdita di infrastrutture critiche, aumentano la probabilità di decisioni errate da parte dei modelli pre-sviluppati. Mentre le grandi imprese o il settore pubblico potrebbero assumere con relativa facilità un maggior numero di avvocati o di informatici per implementare i requisiti per i sistemi ad alto rischio, le aziende più piccole o le start-up non hanno queste possibilità, che devono essere tenute in considerazione.

La posizione negoziale del Consiglio europeo sulla legge sull'AI, adottata il 6 dicembre 2022, è quindi un passo nella direzione sbagliata, in quanto cerca di allentare i requisiti per i sistemi ad alto rischio.[39] Per quanto riguarda la classificazione dei sistemi di AI come ad alto rischio, la proposta di compromesso include ora un livello orizzontale aggiuntivo per garantire che i sistemi di AI che probabilmente non causano gravi violazioni della salute, della sicurezza o dei diritti fondamentali (perché il loro output è solo accessorio a un'azione o a una decisione) siano esentati. Sebbene ciò sia positivo in linea di principio, in quanto alleggerisce l'onere per le start-up innovative, il regolamento orizzontale dovrebbe tenere conto anche del fatto che i rischi previsti cambierebbero in caso di shock esterni o crisi multiple. L'art. L'art. 6(3) della bozza di legge sull'AI dovrebbe quindi essere formulato in modo da esentare quei sistemi in cui il risultato è del tutto irrilevante rispetto all'azione o alla decisione da prendere e, quindi, è improbabile che porti a un rischio significativo per la salute, la sicurezza o i diritti fondamentali, anche nel contesto di una crisi multiple, cioè di shock esterni multipli e simultanei. Per le start-up è particolarmente problematico il fatto che i previsti requisiti orizzontali non saranno emanati prima di un anno dall'entrata in vigore della legge sull'AI.[40] Pertanto, il CEP invita il Parlamento europeo a concentrarsi sui rischi aggiuntivi della previsione algoritmica derivanti dalle crisi e chieda un'attuazione più rapida della norma.

---

[38] The version from 25 November 2022 can be found here: https://data.consilium.europa.eu/doc/document/ST-14954-2022-INIT/de/pdf.Siehe auch: Kullas, M. / Harta, L. (2021), Europäisches Gesetz über Künstliche Intelligenz Kurzfassung, cepAnalyse zu COM2021 206(13.12.2021), https://www.cep.eu/eu-themen/details/cep/europaeisches-gesetz-ueber-kuenstliche-intelligenz-cepanalyse-zu-com2021-206.html.

[39] Stierle, S. (2022), Schwierige Trilog-Verhandlungen im neuen Jahr, Tagesspiegel Background (7.12.2022), https://background.tagesspiegel.de/digitalisierung/schwierige-trilog-verhandlungen-im-neuen-jahr.

[40] Gorzala, J. (2022), AI Act der EU: KI-Regulierung im Anmarsch, Der Brutkasten (7.12.2022), https://brutkasten.com/ai-act-der-eu-ki-regulierung-im-anmarsch/.

Un'altra possibilità è quella di considerare gli effetti delle “policrisi” nella normativa sull'AI nell’ambito della prevista registrazione dei sistemi di AI. Secondo la bozza originale della Commissione, i fornitori di sistemi di AI ad alto rischio, dovrebbero registrare i loro sistemi in un database dell'UE quando entrano nel mercato. Diverse organizzazioni della società civile hanno sottolineato che un regime di trasparenza significativo dovrebbe anche fornire informazioni sull'uso effettivo di questi sistemi nella pratica.[41]Ciò è particolarmente necessario in tempi di crisi multiple, poiché gli esempi sopra descritti dimostrano che i rischi associati alla previsione algoritmica possono essere significativamente più elevati in situazioni anomale e altamente dinamiche rispetto a una valutazione statica del rischio basata sulla mera descrizione dell'attività di una società. Anche in questo caso, obblighi di trasparenza particolarmente stringenti non dovrebbero ostacolare le start-up innovative, ma dovrebbero essere formulati in particolare per quei sistemi di AI utilizzati da imprese dominanti o autorità pubbliche, in quanto questi hanno un impatto potenzialmente più ampio sulla società in caso di un "cigno nero", come l'inizio di una guerra. Pertanto, il CEP accoglie con favore il fatto che la legge sull'AI sia stata rivista per indicare che anche alcuni utenti di sistemi di AI ad alto rischio come autorità pubbliche, istituzioni o altre entità saranno tenuti a registrarsi presso la banca dati UE dei sistemi di AI ad alto rischio.

Nel complesso, è importante sottolineare che le modifiche normative proposte in materia di audit dell'AI e la legge dell'UE sull'AI non intendono limitare l'applicazione dell'AI ad ambienti ben definiti e controllati, in quanto ciò potrebbe ridurre gli effetti cruciali dell'apprendimento. L'obiettivo è piuttosto quello di trovare "regole robuste" che consentano esattamente questo: un'applicazione sicura a un sistema complesso, come è senza dubbio l'attuale crisi multipla. Una volta che i legislatori avranno definito le condizioni quadro normative in modo tale che l'uso di sistemi basati sull'AI per le decisioni economiche o politiche non possa causare danni alla società nel breve termine, il mercato potrà separare gli strumenti di previsione basati sull'AI sensati da quelli non sensati nel lungo termine.

5 Conclusioni: il bisogno di regole solide

Nel complesso, piuttosto che mettere a punto algoritmi per le crisi attuali, i politici e i regolatori europei dovrebbero stabilire regole generali del gioco che, pur non massimizzando l'efficienza, consentano alle economie sempre più guidate dalla tecnologia di operare in modo ragionevole se e quando il mondo entrerà nella prossima crisi.

In generale, l'imprevedibilità degli eventi "cigno nero" e "rinoceronte grigio" suggerisce che un approccio puramente basato sul rischio, come proposto nella legge europea sull'AI, potrebbe non essere sufficiente, poiché non possiamo conoscere il rischio complessivo di un determinato sistema. Tuttavia, se si accetta l'approccio basato sul rischio dell'attuale bozza, si potrebbero incorporare i pericoli che sorgono in tempi di crisi multiple classificando una percentuale maggiore di sistemi guidati dall'AI come "ad alto rischio" ogni volta che l'attuale clima economico o politico suggerisce che i dati potrebbero non essere in linea con la realtà. Soprattutto, i leader politici, gli imprenditori e i giornalisti, che abbracciano con entusiasmo il potenziale degli algoritmi, devono comunicare e comprendere meglio i rischi che ne conseguono in caso di una crisi multipla.

[41] Aszódi, N. (2022), Wie die Regierung bei den Risiken von KI wegschaut, Tagesspiegel Background (6.12.2022), https://background.tagesspiegel.de/digitalisierung/wie-die-regierung-bei-den-risiken-von-ki-wegschaut.

Author:

Dr.Anselm Küsters, LL.M., Head of Division Digitalisation and New Technologies
kuesters@cep.eu

Centrum für Europäische Politik FREIBURG | BERLIN
Kaiser-Joseph-Straße 266 | D-79098 Freiburg
Schiffbauerdamm 40 Raum 4315 | D-10117 Berlin
Tel. + 49 761 38693-0

Traduzione (dalla versione in lingua inglese):
Dott.ssa Eleonora Poli, Responsabile analisi economiche e "business engagement"
poli@cep.eu

Centro Politiche Europee ROMA
Via G. Vico, 1 | I-00196 Roma
Tel. +390684388433
cepitalia@cep.eu

Centrum für Europäische Politik FREIBURG | BERLIN,
Centre de Politique Européenne PARIS
Centro Politiche Europee ROMA
costituiscono il **Centres for European Policy Network** FREIBURG | BERLIN | PARIS | ROMA.

Gli istituti della rete CEP sono specializzati nell'analisi e nella valutazione degli atti promossi dalle istituzioni dell'Unione europea nell'ambito delle politiche di loro competenza e nel quadro d'insieme del processo di integrazione. Il lavoro scientifico, riflesso in particolare nelle proprie pubblicazioni, viene portato avanti indipendentemente da qualsiasi interesse di parte e in favore di una Unione europea che rispetti lo stato di diritto ed i principi dell'economia sociale di mercato.