Osservazioni della presidente von der Leyen al vertice sulla sicurezza dell'Intelligenza Artificiale*

Bletchley Park, Bruxelles, 2 novembre 2023

"Come preparato"

Sessione I: priorità di sicurezza dell'IA per il 2024 e oltre.

Egregio Primo Ministro Sunak, Rishi, Qui a Bletchley Park, onoriamo la grande mente di Alan Turing, il padre dell'informatica moderna. Non so quanto siamo vicini a macchine in grado di ragionare. C'è chi dice che esisterà tra 5 anni, chi no. Ieri se ne è discusso molto.

Stiamo entrando in un'era completamente diversa. Siamo all'alba di un'era in cui le macchine possono agire in modo intelligente. Il mio augurio per i prossimi cinque anni è che impariamo dal passato e agiamo in fretta!

Data la complessità di queste macchine intelligenti, la sicurezza dell'IA è di conseguenza molto più complessa. L'esperienza di altre tecnologie può quindi essere la nostra guida. Prendiamo la storia dell'energia atomica e della bomba nucleare. Gli scienziati hanno scoperto la fisica quantistica che ha portato all'energia nucleare – buona – ma anche con rischi sociali, e anche alla bomba atomica.

Questo insegna una prima importante lezione: l'indipendenza della comunità scientifica è essenziale. Abbiamo bisogno di un sistema di pesi e contrappesi scientifici oggettivi. Abbiamo bisogno di coltivare una comunità di scienziati eccezionali e indipendenti. Scienziati, con accesso alle risorse per valutare i rischi dell'IA e liberi di denunciare tali rischi.

In secondo luogo, dobbiamo stabilire standard di sicurezza dell'IA che siano accettati in tutto il mondo.

L'aviazione è una buona fonte di ispirazione. Viaggiare in aereo è diventato estremamente sicuro perché abbiamo imparato sistematicamente dagli errori. Qualsiasi errore può portare a un risultato catastrofico. La comunità dell'aviazione ha reso una pratica standard che qualsiasi incidente o errore sia reso pubblico e seguito. Non è visto come un fallimento, ma come responsabile e appropriato segnalare un errore. Qualsiasi errore viene analizzato e i risultati sono disponibili al pubblico e le raccomandazioni

* traduzione redazionale, non ufficiale e non riveduta

sono sempre seguite. Questo approccio dimostra il valore di standard e procedure condivise.

In terzo luogo, possiamo anche imparare dall'esperienza più recente nello sviluppo di una cultura della sicurezza informatica. Le organizzazioni sono meglio preparate contro gli attacchi informatici quando c'è un'efficace condivisione delle informazioni. Gli avvisi di sicurezza ampiamente condivisi prevengono la peggiore diffusione virale. E' una questione di tempo. Anche i sistemi di intelligenza artificiale si evolvono e imparano. Gli algoritmi complessi non possono mai essere testati in modo esaustivo. Quindi, prima di ogni altra cosa, dobbiamo assicurarci che gli sviluppatori agiscano rapidamente quando si verificano problemi, sia prima che dopo che i loro modelli sono stati immessi sul mercato. In breve, spero che tra 5 anni avremo tutti sistemi in grado di mettere in atto queste lezioni. Fare tutto questo è la chiave per sbloccare gli enormi vantaggi dell'intelligenza artificiale.

Sessione II: Passi concreti per rendere sicura l'IA di frontiera

L'estate scorsa, ho detto che dovremmo unire le forze per un approccio globale alla comprensione dell'impatto dell'IA. A quel tempo, ho pensato a qualcosa come l'IPCC per il clima. Da allora, il dibattito è stato così intenso, ad esempio nel nostro processo del G7 di Hiroshima, che vediamo ancora più chiaramente ciò che è necessario.

Credo che un quadro per comprendere e mitigare i rischi di sistemi di IA molto complessi dovrebbe essere costruito su 4 pilastri.

E ne abbiamo discusso nella prima sessione:

in primo luogo, abbiamo bisogno di una comunità scientifica fiorente e indipendente, dotata dei mezzi per valutare i sistemi di intelligenza artificiale. Ha bisogno di finanziamenti pubblici e dell'accesso ai migliori supercomputer. Negli ultimi 5 anni l'UE ha costruito la più grande rete pubblica di supercomputer al mondo. E già diamo accesso a Lumi in Finlandia e Leonardo in Italia a start-up e tester.

In secondo luogo, dobbiamo sviluppare procedure e standard accettati a livello internazionale per testare la sicurezza dell'IA.

In terzo luogo, deve essere una procedura standard, che ogni incidente significativo causato da errori o uso improprio dell'IA sia segnalato e seguito.

Quarto, abbiamo bisogno di un sistema internazionale di allerta alimentato da segnalatori attendibili.

Questi 4 pilastri dovrebbero costituire un sistema di governance efficace.

Ora, al centro di tutto, c'è bisogno di una cultura della responsabilità.

Per gli attori privati, questo significa un principio generale: maggiore è la capacità del modello di IA e i rischi che ne derivano, maggiore è la responsabilità. Ciò significa una responsabilità aziendale solida e tracciabile, integrata nel proprio modello di business. Ma questo va oltre la pura responsabilità aziendale.

Le autorità pubbliche sono responsabili in ultima istanza della sicurezza e dell'incolumità dei cittadini. Dobbiamo quindi mettere in atto norme vincolanti di principio. E le autorità pubbliche devono avere poteri di intervento, come complemento e sostegno all'autoregolamentazione. Questi sono i guardrail.

E proprio come in autostrada: i guardrail non sono barriere, ma consentono al traffico di rimanere sulla strada e di procedere in sicurezza.

Questo è il motivo per cui abbiamo proposto una legge sull'IA. I suoi principi di base consistono nel sostenere l'innovazione, sfruttare i vantaggi dell'IA e concentrare la regolamentazione solo sui rischi elevati.

La legge sull'IA è nelle fasi finali del processo legislativo. In questo processo, stiamo discutendo la creazione di un Ufficio europeo per l'IA. Questo Ufficio potrebbe occuparsi dei modelli di IA più avanzati, con responsabilità di supervisione, nella logica del quadro dei 4 pilastri che ho delineato. L'Ufficio europeo per l'IA dovrebbe collaborare con la comunità scientifica in generale. Potrebbe contribuire a promuovere norme e pratiche di test per i sistemi di IA di frontiera. Potrebbe integrare il settore privato nelle indagini e nei test. Dovrebbe essere in grado di agire sulla base degli avvisi e assicurarsi che gli sviluppatori si assumano la responsabilità.

Infine, un Ufficio europeo per l'IA applicherebbe le norme comuni in tutti i 27 Stati membri per i modelli più avanzati.

Questa è una buona notizia per queste imprese e per la sicurezza in Europa. Ma l'Ufficio europeo per l'IA dovrebbe avere anche una vocazione globale. Dovrebbe essere aperta a cooperare con entità simili in tutto il mondo. Compresi, ovviamente, il caro Rishi, la cara Kamala con i vostri nuovi Istituti per la Sicurezza dell'IA.

Cari colleghi, la storia ci sta guardando. Come disse Mark Twain: "Il segreto per andare avanti, è iniziare".