

Normative e standard su AI

Il quadro internazionale



Matteo Lucchetti

Direttore Operativo, Cyber 4.0

Matteo.Lucchetti@cyber40.it

14/05/2024

Vantaggi dell'AI

- **Benefici dell'AI per le persone**

- Miglioramento dell'assistenza sanitaria
- Veicoli e altri sistemi di trasporto più sicuri
- Prodotti e servizi personalizzati, più economici e duraturi
- Migliore accesso a informazioni, istruzione e formazione
- Sicurezza del luogo di lavoro se i robot sono utilizzati per le azioni potenzialmente più pericolose
- Nuove posizioni lavorative man mano che le industrie guidate dall'AI crescono e cambiano

- **Servizi pubblici**

- Trasporti pubblici, istruzione, gestione dell'energia e dei rifiuti

- **Opportunità per le imprese**

- Sviluppo di nuove generazioni di prodotti e servizi, incremento della produzione e della qualità
- Potenziamento delle vendite, miglioramento e manutenzione delle macchine
- Ottimizzazione energetica

- **Difesa e Sicurezza**

- Supporto a strategie di difesa e attacco, sia sul campo che nello spazio cyber
- Prevenzione dei crimini e sistema di giustizia penale

Artificial Intelligence e cyber security

Cybersecurity difensiva



Malware Detection



Antispam



Data Classification



Vulnerability
Management



Threat Intelligence

Cybersecurity offensiva



Malware Creation



Smart Botnets



Spear Phishing



Conditional Attacks

AI, LLM e rischi di cybersecurity

Scam email cyber attacks increase after rise of ChatGPT

By Marcus Law

April 03, 2023 • 5 mins

+135%

New AI Tool 'FraudGPT' Emerges, Tailored for Sophisticated Attacks

📅 Jul 26, 2023 👤 THN

Cyber Crime / Artificial Intelligence

Business

'Tsunami' of AI-powered phishing, scams forecast in 2024

Arthur Fuentes, ABS-CBN News

Published Jan 23, 2024 05:58 PM PHT



Deepfake e operazioni cyber

Forbes

Sep 3, 2019, 04:42pm EDT | 47.443 views

A Voice Deepfake To Scam A CEO \$243,000



Jesse Damiani Contributor ⓘ

Consumer Tech

I cover the human side of VR/AR, Blockchain, and Media.

Microsoft's VALL-E Generates Speech From Just 3 Seconds of Audio

But it could lead to a proliferation of deepfake voices



Ben Wodecki

January 11, 2023

🕒 2 Min Read



Notification

INVESTIGATION, CYBER DIVISION

Information is being provided by the FBI, with no guarantee, for potential use at the sole discretion of the recipient against cyber threats. This data is provided to professionals and system administrators guard against malicious actions of cyber actors. This PIN was 1234567890.

Classified **TLP:WHITE**: Subject to standard copyright and trademark laws, information may be distributed without restriction.

Deepfakes Almost Certainly Will Be Used to Create Synthetic Content for Cyber and Other Operations

Deepfake



Deepfake per motivazioni politiche

NEWS

Deepfake video of Zelensky telling Ukrainians to surrender removed from social platforms

By Joshua Rhett Miller

March 17, 2022 | 12:20pm | Updated



CYBER ATTACK

Iran-backed hackers interrupt UAE, UK and Canadian programming with fake AI news broadcast

A group of hackers linked to Iran have interrupted BBC and a host of other European TV streaming services in Britain, the United Arab Emirates and Canada, Microsoft stated in a report earlier this month, noting a marked acceleration of Iranian cyber attacks since Hamas's October 7 attack on Israel. The programming was interrupted with a fake news report on Gaza featuring graphic images and what appeared to be an AI-generated anchor – the first time Iran has used AI in this way in its influence operations.

Issued on: 14/02/2024 - 16:09 ⌚ 4 min



L'arresto di Trump



Allucinazioni dell'AI



Impatto dell'utilizzo di applicazioni AI sui diritti fondamentali

- **Bias e discriminazione**

- Impatti documentati su discriminazione di genere, razziale, sociale.

- **Autonomia dell'individuo, libertà di espressione e di informazione, diritto ad elezioni libere**

- Es. campagne di disinformazione, filter bubbles, manipolazione, rimozione di contenuti specifici, Fake news.

- **Protezione dei dati personali e diritto alla privacy**

- Es. Uso di sistemi di identificazione biometrica basati su AI

- AI e sistemi automatici di decision-making → **buona amministrazione, accesso alla giustizia, diritto a un giusto processo**

- Es. Policy di prevenzione (+50% delle forze di polizia in UE), modelli di analisi del rischio, social scoring

Perché regolamentare l'AI

- **Protezione dei Dati e della Privacy:**

- La regolamentazione è fondamentale per garantire che l'uso dell'AI rispetti la privacy e protegga i dati personali degli individui, in linea con standard come il GDPR.

- **Sicurezza e Affidabilità**

- Leggi specifiche sull'AI possono stabilire **standard di sicurezza per prevenire malfunzionamenti o abusi** che potrebbero avere gravi conseguenze.

- **Etica e Diritti Umani**

- Una legislazione può assicurare che l'AI venga sviluppata ed utilizzata in modo etico, **rispettando i diritti umani e evitando discriminazioni**.

- **Trasparenza e Responsabilità**

- Le leggi possono imporre ai creatori e agli utilizzatori di AI di essere **trasparenti sulle modalità di funzionamento** e di essere **responsabili per le decisioni prese dagli algoritmi**.

- **Promozione dell'Innovazione**

- Una chiara regolamentazione può promuovere l'innovazione, fornendo un quadro sicuro e stabile per la **ricerca** e lo **sviluppo**.

- **Competitività Economica**

- Legiferare sull'AI può aiutare a posizionare un paese come leader nel campo dell'innovazione tecnologica, attrarre **investimenti** e stimolare la **crescita economica**.

- **Governance e Supervisione**

- Una legge può istituire **organismi di governance** per monitorare l'evoluzione dell'AI e intervenire in caso di necessità

UN security council delegates urge AI controls to defuse potential global threat



**United
Nations**

Meetings Coverage and Press Releases

**General Assembly Adopts Landmark Resolution on Steering
Artificial Intelligence towards Global Good, Faster Realization
of Sustainable Development**

La Risoluzione ONU del 21/3/2024

«Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development»

- La risoluzione mira a un consenso globale su sistemi di AI sicuri e affidabili, promuovendo **approcci normativi nazionali** e sostenendo lo sviluppo sostenibile.
- **Non fornisce una definizione di AI**, ma si riferisce a sistemi non militari, centrati sull'uomo e rispettosi dei diritti umani.
- La risoluzione **non ha uno scopo territoriale o settoriale specifico**, ma è globale e intersettoriale, senza ruoli di conformità espliciti.
- Cerca di colmare il divario digitale e di promuovere sistemi di AI sicuri, sostenendo i paesi in via di sviluppo verso la trasformazione digitale e sottolineando la **governance dei dati** e il **rispetto dei diritti umani**.
- **Non impone requisiti di conformità**, ma invita all'azione per l'accesso ai benefici della trasformazione digitale e allo sviluppo sostenibile.
- **Non è vincolante legalmente**

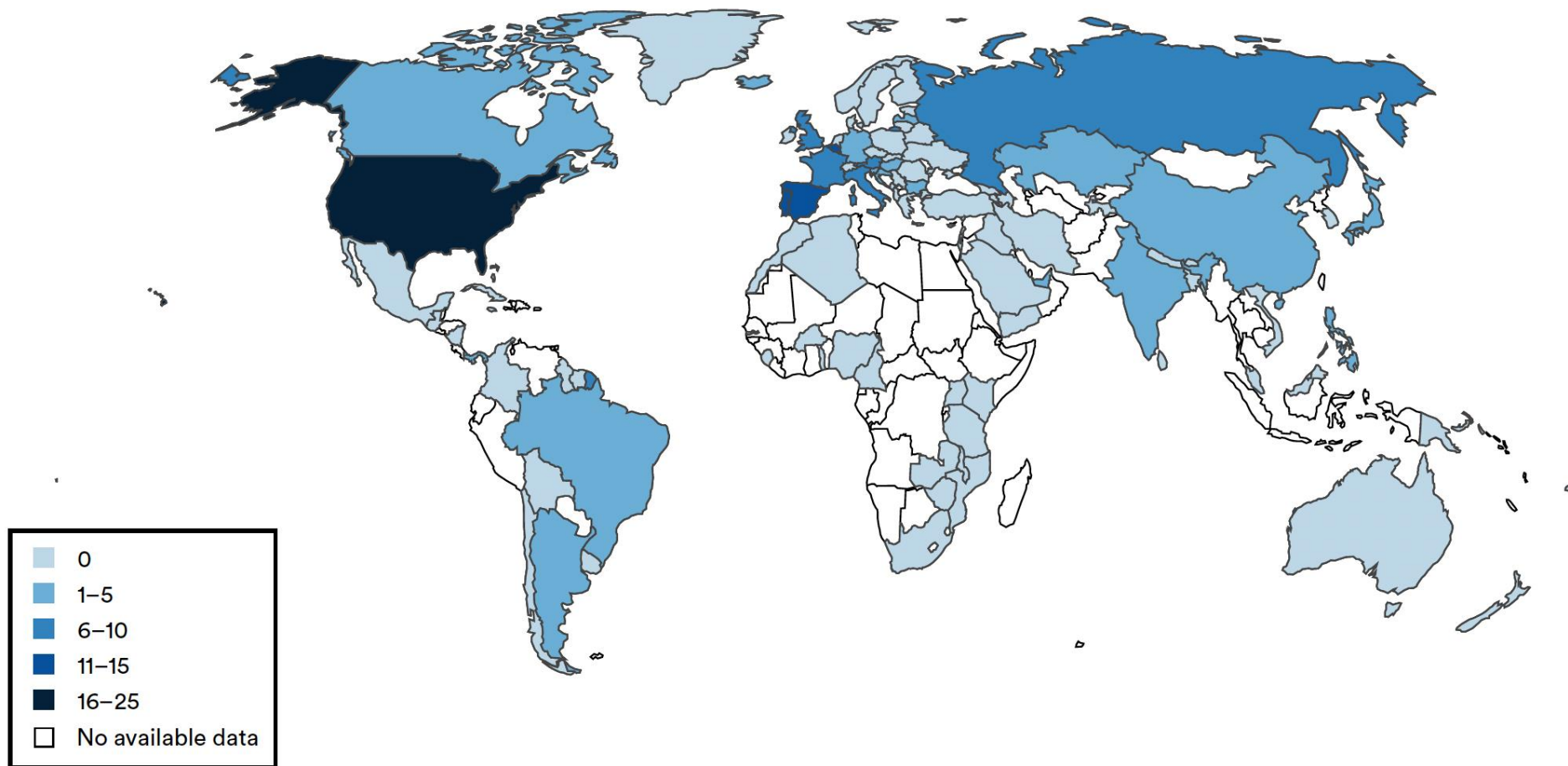
Altri strumenti di organizzazioni sovranazionali

- **OCSE**
 - AI Principles
 - AI Risk Classification Framework
- **UNESCO**, AI Ethics Recommendations
- **Consiglio d'Europa**, Framework Convention on AI, Human Rights and Rule of Law
- **Unione Europea**, AI Act

Normative con componenti relative all'AI – Contesto internazionale

Number of AI-related bills passed into law by country, 2016–23

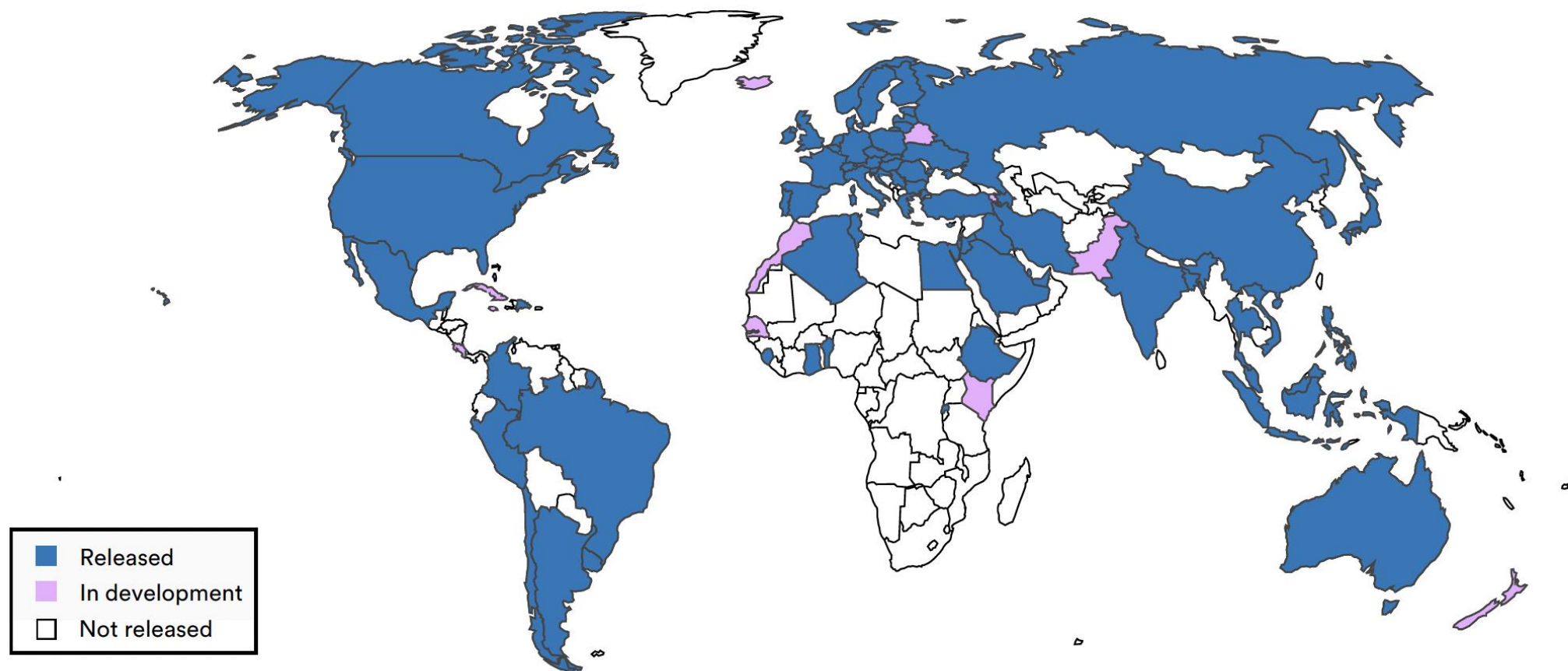
Source: AI Index, 2024 | Chart: 2024 AI Index report



Strategie AI – Contesto internazionale

Countries with a national strategy on AI, 2023

Source: AI Index, 2024 | Chart: 2024 AI Index report



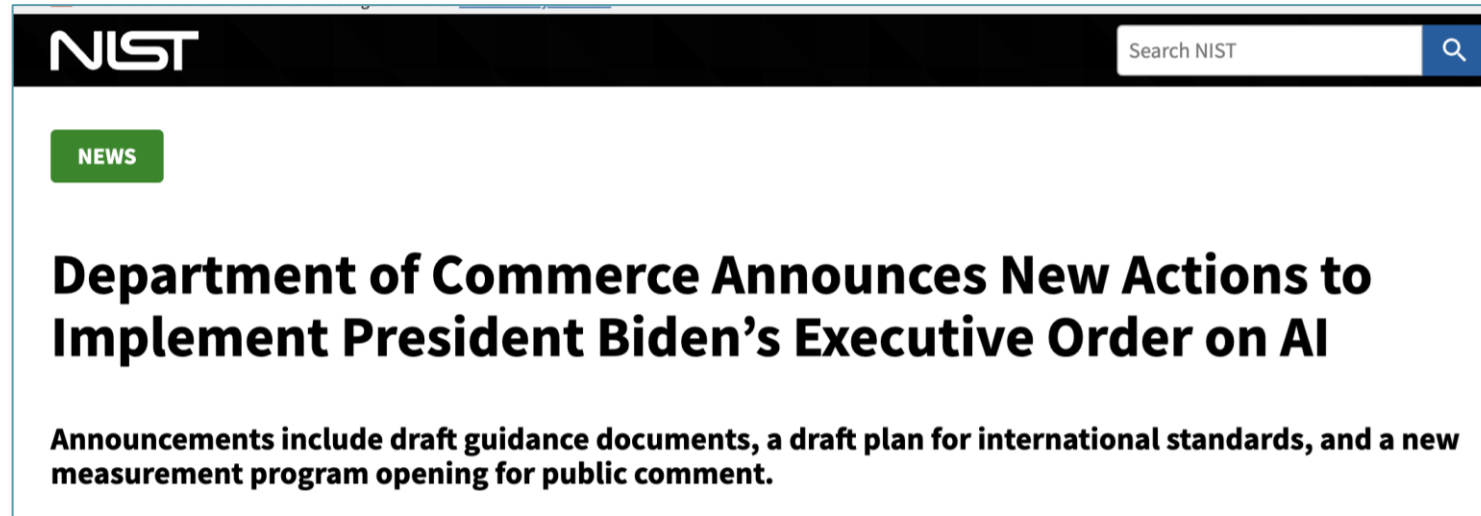
- **Ordine Esecutivo della Casa Bianca sull'AI** (Sviluppo e Uso Sicuro, Protetto e Affidabile dell'Intelligenza Artificiale)

- "Sfruttare l'AI per il bene e realizzare i suoi innumerevoli benefici richiede la mitigazione dei suoi sostanziali rischi."
- Numerosi settori coinvolti, agenzie federali e sviluppatori di soluzioni AI
- Sviluppo di standard federali
- Sviluppatori dei sistemi di AI più potenti devono condividere i risultati dei test di sicurezza e altre informazioni critiche con il governo
- Linee guida per l'autenticazione dei contenuti e il watermarking per etichettare i contenuti generati dall'AI

- **Blueprint della Casa Bianca per una Carta dei Diritti dell'AI**

- Linee guida per accesso e utilizzo equo dei sistemi AI, cinque principi e pratiche associate per guidare la progettazione, l'uso e la diffusione di "sistemi automatizzati": Safe and Effective Systems, Algorithmic Discrimination Protections, Data Privacy, Notice and Explanation, and Human Alternatives, Consideration, and Fallback
- Diverse aziende leader nell'AI (ad esempio, Adobe, Amazon, IBM, Google, Meta, Microsoft, Open AI e Salesforce) si sono impegnate **volontariamente** a: test di sicurezza interni/esterni dei sistemi di AI prima del rilascio, condivisione di informazioni sulla gestione dei rischi, investimento in misure di sicurezza.

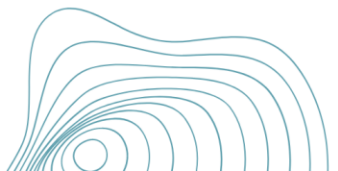




- *AI RMF Generative AI Profile* ([NIST AI 600-1](#)) – Identificazione e gestione dei rischi associati all'AI generativa
- *Secure Software Development Practices for Generative AI and Dual-Use Foundation Models* (NIST [Special Publication \(SP\) 800-218A](#))
- *Reducing Risks Posed by Synthetic Content* ([NIST AI 100-4](#)) – Approcci tecnici per promuovere la trasparenza nei contenuti digitali
- *A Plan for Global Engagement on AI Standards* ([NIST AI 100-5](#)) – Promozione e sviluppo del consenso globale su standard AI, cooperazione e coordinamento, condivisione di informazioni

Interim Measures for the Management of Generative Artificial Intelligence Services

- Entrate in vigore il 15 Agosto 2023
- Non stabiliscono categorie di rischio per i servizi di AI, anche se alcune disposizioni suggeriscono che alcune categorie di servizi di AI sono soggette a un controllo più rigoroso da parte dei regolatori
- I fornitori di servizi di AI generativa con “attributi di opinione pubblica o la capacità di mobilitazione sociale” sono tenuti a effettuare una valutazione della sicurezza e a sottoporsi a una verifica normativa prima di lanciare tali servizi, oltre che rispettare i requisiti generalmente applicabili come la moderazione dei contenuti e l’etichettatura



- **Utilizzo lecito:** Elaborazione dei dati in modo lecito, rispetto dei diritti di proprietà intellettuale e ottenimento del consenso per l'uso delle informazioni personali
- **Regole di etichettatura dei dati:** Stabilire regole di etichettatura chiare, garantire un'etichettatura accurata e fornire formazione al personale addetto all'etichettatura. Inoltre, le Misure AI richiedono specificamente che un fornitore di servizi di AI generativa rispetti quanto segue:
 - **"Sostenere i valori fondamentali socialisti"** – incluso il non generare contenuti proibiti, come contenuti che incitano alla **"sovversione del potere statale o al rovesciamento del sistema socialista"**, mettono in pericolo la sicurezza nazionale e gli interessi, danneggiano

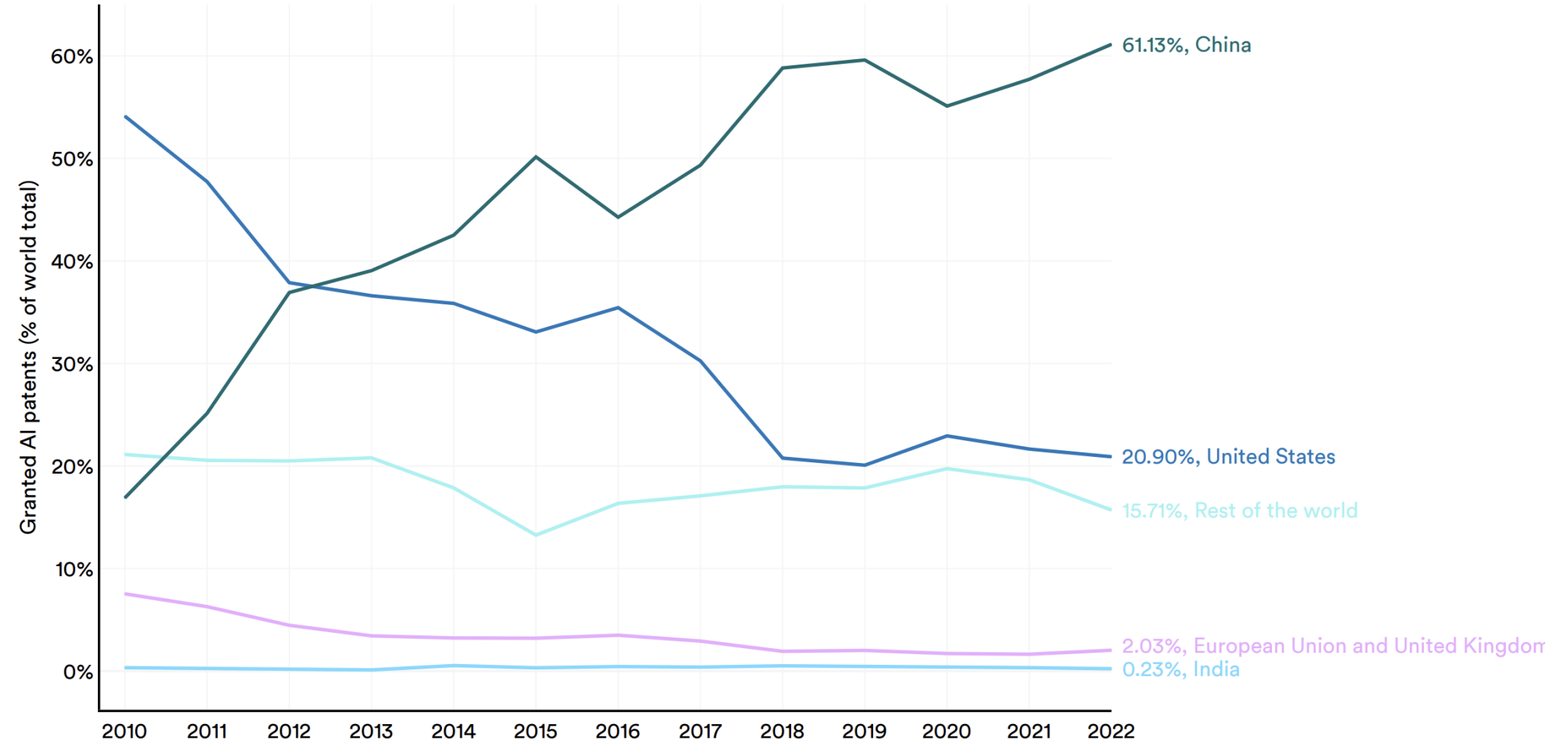
l'immagine nazionale, incitano alla divisione del paese, minano l'unità nazionale e la stabilità sociale, fanno apologia del terrorismo, estremismo, odio etnico e discriminazione, violenza, pornografia, e informazioni false e dannose".

- "prevenire la discriminazione basata su nazionalità, religione, paese, regione, genere, occupazione e salute nella progettazione dell'algoritmo, nel set di dati di addestramento, nella generazione del modello, nell'ottimizzazione e nella fornitura del servizio".
- Rispetto dei diritti di proprietà intellettuale e dell'etica commerciale.
- Rispetto dei legittimi diritti e interessi altrui.
- "migliorare la trasparenza, l'accuratezza e l'affidabilità dei contenuti generati dai servizi di AI generativa".

Standard tecnici internazionali

- ISO/IEC 22989:2022 – Information technology – Artificial intelligence – **Artificial intelligence concepts and terminology**
- ISO/IEC 23053:2022 – **Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML)**
- **ISO/IEC 42001:2023** – Information technology – Artificial intelligence – **Management Systems**
 - **39 controlli su 9 clausole**
 - Politiche relative all'AI
 - Organizzazione interna (es. ruoli e responsabilità).
 - Risorse per i sistemi di AI (es. dati, strumenti, risorse umane).
 - Analisi dell'impatto dei sistemi di AI (es. su individui, gruppi e società e relativa documentazione).
 - Ciclo di vita del sistema AI.
 - Dati per i sistemi di AI (es. acquisizione e preparazione dei dati).
 - Informazioni per le parti interessate ai sistemi di AI (es. comunicazione di incidenti).
 - Utilizzo dei sistemi di AI (es. uso responsabile ed uso previsto).
 - Rapporti con terze parti (ad esempio: fornitori, clienti)

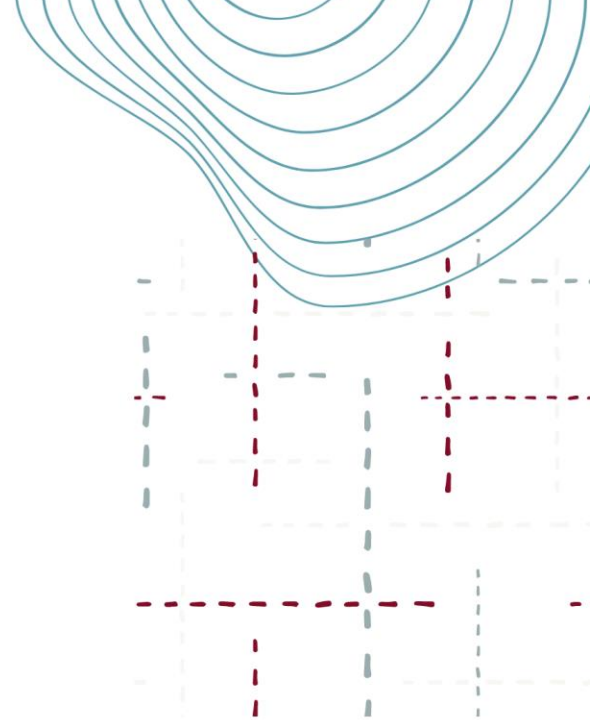
Source: Center for Security and Emerging Technology, 2023 | Chart: 2024 AI Index report





Grazie a tutti

14/05/2024



Matteo Lucchetti
Direttore Operativo, Cyber 4.0
Matteo.Lucchetti@cyber40.it