NIST AI Risk Management Framework, una risorsa per le organizzazioni



Poco più di un anno fa, nel gennaio del 2023, il National Institute of Standards and Technology (NIST) ha diffuso un documento dal titolo *"Artificial Intelligence Risk Management Framework – AI RMF 1.0"* al fine di contribuire alla mitigazione dei rischi legati all'Intelligenza Artificiale.

Il documento è accompagnato dall'*AI RMF Playbook*, che fornisce ulteriori strumenti per l'implementazione delle misure raccomandate dal Framework.

La strategia USA sull'Intelligenza Artificiale

L'AI RMF scaturisce da un'elaborazione triennale che ha visto l'agenzia statunitense collaborare con diversi partner pubblici e privati.

Le fitte consultazioni con gli esponenti delle *Big Tech* caratterizzano da diversi anni i lavori del governo di Washington sul tema; in questo parzialmente differenziandosi dalla strategia europea, che vede un minor contributo dei privati alla nascente regolamentazione di settore.

Nella medesima direzione andavano, d'altronde, tanto l'introduzione del Chief Digital and Artificial Intelligence Office (CDAO) presso il Dipartimento della Difesa USA nel 2022 quanto l'*Executive Order on Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence* emesso dall'amministrazione Biden-Harris l'anno successivo.

AI RMF 1.0: struttura e contenuti

L'*Artificial Intelligence Risk Management Framework* si articola in due parti, la prima delle quali ("***Foundational Information***") è suddivisa in 4 sotto-sezioni:

- ***Framing Risk.*** Dopo aver definito il rischio come *"composite measure of an event's probability of occurring and the magnitude or degree of the consequences"*, il NIST riconosce che la valutazione quantitativa/qualitativa del rischio legato agli strumenti AI, soprattutto nelle prime fasi del loro ciclo di vita, presenta delle difficoltà; sottolineando come la pretesa di "azzerare" i livelli di rischio possa addirittura risultare "controproduttiva", dovendosi pertanto preferire misure e strategie mirate alla loro ragionevole mitigazione.
- ***Audience.*** Ritenendo che ogni strategia di *risk management* dipenda dalla *"collective responsibility"* il documento evidenzia la necessità che i relativi processi abbiano uno stampo multidisciplinare, includendo diverse professionalità, prospettive e discipline per garantirne l'efficacia e il successo.
- ***AI Risks and Trustworthiness.*** Rispetto all'affidabilità dei sistemi basati sull'Intelligenza Artificiale, il Framework cita diverse caratteristiche che andrebbero sempre contemplate (e bilanciate fra loro) nella progettazione degli stessi:

1. *safe,*
2. *valid and reliable,*
3. *secure and resilient,*
4. *accountable and transparent,*
5. *explainable and interpretable,*
6. *privacy-enhanced,*
7. *fair with harmful bias managed.*

- ***Effective.*** Come già avvenuto per la stesura del

Framework, rispetto alla valutazione della sua efficacia e agli eventuali aggiornamenti futuri il NIST dichiara l'intenzione di voler procedere *"in conjunction with the AI community".*

Anche la seconda parte (dal titolo *"Core and Profiles"*) si articola in quattro sezioni, corrispondenti ad altrettante funzioni essenziali per la gestione dei rischi connessi all'impiego dell'Intelligenza Artificiale:

- ***Govern.*** Rispetto al ruolo delle istituzioni governative il documento cita come esempi: *"policies, processes, and procedures for AI systems; empowering teams responsible for AI systems; promoting workforce diversity, equity, inclusion, and accessibility; and implementing processes for robust engagement with relevant AI actors".*
- ***Map.*** L'importanza di questa funzione deriva dalla complessità dell'*AI system lifecycle,* che impone di coordinare diverse prospettive e numerosi attori al fine di *"develop more trustworthy AI systems by improving their capacity for understanding contexts or anticipating risks of non-intended uses of AI systems"*.
- ***Measure.*** La funzione di "misurazione" e monitoraggio dei pericoli legati all'AI e dei relativi impatti, finalizzata a fornire "a traceable basis to inform management decisions", include la valutazione dei sistemi in base alle loro caratteristiche (specificate nella Parte I), oltre a meccanismi per tracciare l'evoluzione dei rischi nel tempo e a sistemi di feedback circa la validità e attualità delle misurazioni stesse.
- ***Manage.*** L'ultima funzione raggruppa tutte le attività associate alla gestione operativa dei rischi (anche relativi a *"third-party AI systems"*) identificati lungo i passaggi precedenti: ciò include la loro classificazione in ordine di priorità e la definizione di strategie per *"maximize AI benefits and minimize*

negative impacts” nonché la scelta delle più idonee contromisure, che devono sempre includere *“response and recovery plans”.*

Quanto infine ai destinatari cui si rivolge, il Framework dichiara l’ambizioso obiettivo di rappresentare una risorsa per ogni individuo, organizzazione o società – a prescindere dal settore di riferimento – le cui operazioni includano attività di *“designing, developing, deploying or using AI systems, to help manage the many risks of AI and promote trustworthy and responsible development and use of AI systems”.*

A cura della Redazione

Condividi sui Social Network:
https://www.ictsecuritymagazine.com/notizie/nist-ai-risk-management-framework-una-risorsa-per-le-organizzazioni/