

QUADERNI DI ReD OPEN

a cura del ReD OPEN Competence Center

# VALUTARE L'INTELLIGENZA ARTIFICIALE

Una panoramica introduttiva su norme  
e valutazioni d'impatto dell'IA



QUADERNI DI ReD OPEN

# **VALUTARE L'INTELLIGENZA ARTIFICIALE**

**Una panoramica introduttiva  
su norme e valutazioni d'impatto  
dell'IA**

**a cura del ReD OPEN Competence Center**

**Ledizioni**



Attribuzione - Non commerciale - Non opere derivate 4.0  
Internazionale (CC BY-NC-ND 4.0)

2024 Ledizioni LediPublishing  
Via Boselli 10, 20136 Milano  
<http://www.ledizioni.it>  
e-mail: [info@ledizioni.it](mailto:info@ledizioni.it)

Prima edizione Ledizioni: giugno 2024

*Valutare l'Intelligenza Artificiale. Una panoramica introduttiva  
su norme e valutazioni d'impatto dell'IA*

ISBN cartaceo 9791256001606  
ISBN ePub 9791256001613  
ISBN PDF 9791256001620

Tutti i marchi esposti e citati in questa pubblicazione sono di proprietà dei rispettivi detentori dei copyright; marchi di terzi, nomi di prodotti, nomi commerciali, nomi corporativi e società citati possono essere marchi di proprietà dei rispettivi titolari o marchi registrati d'altre società e appartengono ai loro legittimi proprietari.

In copertina: image cover by Freepik.com

Informazioni sul catalogo e sulle ristampe: [www.ledizioni.it](http://www.ledizioni.it)

# INDICE

## PARTE I. PERCHÉ SI PARLA DELLE VALUTAZIONI D'IMPATTO DELL'INTELLIGENZA ARTIFICIALE

<b>Introduzione</b>	<b>15</b>
L'approccio basato sul rischio al centro del Regolamento europeo sull'IA	20
Panoramica delle valutazioni d'impatto	23
Cosa sono le valutazioni dell'impatto algoritmico?	24
Principali fasi di AIA	26

## PARTE II. GLI STRUMENTI

<b>Introduzione</b>	<b>31</b>
Canadian Algorithmic Impact Assessment (AIA) tool (Governo del Canada)	33
Algorithmic Impact Assessment (AIA) (Federal CIO Council)	35
AI systems ethics self-assessment (Digital Dubai Authority)	37
From Principles to Practice: An interdisciplinary framework to operationalize AI ethics (AI Ethics Impact Group)	39
Artificial Intelligence Impact Assessment (Considerati / ECP)	41
Unlocking Public Sector AI. AI Procurement in a Box (World Economic Forum)	43
Automated Decision Impact Assessment (ADIA) Framework (Open Loop/Facebook)	45

Automated Decision Making Systems in the Public Sector. An Impact Assessment Tool for Public Authorities (Algorithm Watch)	47
Algorithmic impact assessment (Ada Lovelace Institute)	49
Cap AI- A procedure for conducting conformity assessment of AI systems in line with the EU Artificial Intelligence Act (Oxford University)	51
Audit Framework for Algorithms (Netherlands Court of Audit)	53
Model rules on impact assessment of automated decision-making systems used by public administration (European Law Institute)	56
Fundamental rights and algorithm impact assessment - FRAIA (Netherlands Court of Audit)	58
Ethical Impact Assessment (UNESCO)	61
AI Impact Assessment: a tool to set up responsible AI projects	64
AI Governance Framework for Generative AI (Singapore)	67
Artificial Intelligence Risk Management Framework (NIST)	71

### PARTE III. RIFLESSIONI ED ESPERIENZE

<b>Riflessioni: Human oversight</b>	<b>77</b>
<b>Esperienze: l'approccio di Haier Europe per un uso responsabile dell'Intelligenza Artificiale</b>	<b>83</b>

### PARTE IV. CONCLUSIONI

<b>Oltre una visione <i>tecnocentrica</i></b>	<b>89</b>
<b>L'approccio olistico alla governance dell'IA</b>	<b>97</b>
<b>Postfazione</b>	<b>99</b>

<b>AI DIGITAL SONAR</b>	<b>103</b>
<b>Riferimenti bibliografici</b>	<b>105</b>
<b>Elenco delle valutazioni dell'impatto algoritmico descritte</b>	<b>108</b>





QUADERNI DI RED OPEN

*I principi generali del trattamento dei dati sanitari*

*Governare l'Intelligenza Artificiale*

*E liberaci dal malware*

*La trasformazione digitale del trattamento sanitario*



Questo volume rappresenta l'esordio di AI Digital Sonar, un'iniziativa di ReD Open, spin-off dell'Università degli studi di Milano Bicocca, che ha l'ambizione di favorire la diffusione di un approccio responsabile "by design" all'innovazione tecnologica digitale. L'obiettivo esplicito del progetto è quello di condividere pratiche e riflessioni per identificare, tra percorsi e riflessioni diverse, un approccio responsabile alla gestione by design per la protezione dei dati e per l'adozione di tecnologie di IA. Le valutazioni d'impatto raccolte ed elencate in questo breve volume, vogliono essere uno stimolo per tutti noi per prendere coscienza della situazione attuale, ossia per conoscere, comprendere e governare i potenziali impatti trasformativi degli strumenti che utilizzano l'IA. Durante il 2024 ci saranno occasioni di aggiornamento sullo stato delle valutazioni d'impatto, nonché ulteriori iniziative legate al progetto AI DIGITAL SONAR.

Per rimanere aggiornati:



AI DIGITAL SONAR ad oggi ha ricevuto il sostegno di:

**Haier Europe**



**AG** | arlatighislandi  
CONSULENZA DEL LAVORO E FISCALE



## PARTE I

# PERCHÉ SI PARLA DELLE VALUTAZIONI D'IMPATTO DELL'INTELLIGENZA ARTIFICIALE



## INTRODUZIONE

Il dibattito sulle tecnologie di Intelligenza Artificiale (IA) e, in particolare, su come governare il loro irrompere nei più svariati ambiti e livelli organizzativi non può prescindere dal concetto di accountability. A dispetto della sua complessità – che ha portato alcuni autori a parlare espressamente di “bizantinismo” – l’accountability si configura come un “concetto ombrello”, dotato di grande potenziale esplicativo. Negli ultimi anni, sono state proposte numerose definizioni di tale nozione. La maggior parte di queste si richiama all’idea di: “Limitare o vincolare l’insieme di azioni alternative che un attore è autorizzato a considerare prima di decidere cosa fare o cosa dire”. Gli individui sono chiamati, pertanto, a spiegare e, in alcuni casi, a giustificare i loro comportamenti e le loro decisioni (o anche non decisioni), per non incorrere in sanzioni, sia di natura formale sia informale. La letteratura su questi temi concepisce l’accountability come: “Una relazione sociale in cui un attore si sente obbligato a spiegare e giustificare il proprio comportamento a un altro attore significativo”. In quest’ottica, Mark Bovens identifica cinque tipologie di accountability, rispetto alla natura del forum, ossia ai soggetti o alle entità verso cui si “rende conto”: professional accountability; political accountability; administrative accountability; legal accountability; social accountability. Ai nostri giorni è lecito domandarsi se gli strumenti di intelligenza artificiale siano o meno accountable. Sicuramente, a segui-

to dell'entrata in vigore dell'AI Act potranno essere accountable da un punto di vista legale. Ma potranno, invece, diventare accountable anche nei confronti della società, intesa nel suo insieme, ovvero delle persone che la compongono e che spesso, non avendo competenze per leggerne criticamente il funzionamento, utilizzano questi strumenti in modo inconsapevole?

È lecito dunque, e anzi doveroso, porsi questa domanda ora che i sistemi basati sull'intelligenza artificiale (IA) sono al centro dell'interesse di istituzioni, organizzazioni e individui per il potenziale di miglioramento dei processi, dei prodotti e dei servizi che promettono, e che in parte già realizzano. L'IA è passata in pochi anni da argomento di nicchia a tema centrale del dibattito pubblico, delle strategie aziendali – non solo delle imprese di grandi dimensioni – e delle discussioni che hanno luogo all'interno delle arene politiche, ai diversi livelli di governance. Questa trasformazione risulta oltremodo evidente nello scenario europeo.

In particolare, l'IA negli ultimi anni ha sempre più interessato il panorama giuridico europeo. A partire dal 2018, infatti, il dibattito sull'IA si è concretizzato nelle istituzioni europee, con un momento chiave rappresentato dalla Comunicazione della Commissione Europea intitolata “L'Intelligenza Artificiale per l'Europa”, che ha introdotto per la prima volta una definizione ufficiale di IA come sistemi capaci di comportamento intelligente attraverso l'analisi dell'ambiente e l'azione autonoma per raggiungere obiettivi specifici. In questo contesto, è stato creato il Gruppo di esperti di alto livello sull'intelligenza artificiale (*High-Level*



*Expert Group on Artificial Intelligence*, AI HLEG) per supportare l'attuazione della strategia europea sull'IA, orientata verso lo sviluppo di un'intelligenza artificiale affidabile. Nel 2019, il gruppo ha presentato le *Ethics Guidelines for Trustworthy AI*, un insieme di linee guida etiche basate sui diritti fondamentali definiti dai Trattati europei e dalla Carta dei diritti fondamentali dell'UE, che includono il rispetto della dignità umana, la libertà individuale, la democrazia, la giustizia, lo stato di diritto, l'uguaglianza, la non discriminazione, la solidarietà, e i diritti delle minoranze, nonché il diritto all'informazione. Queste linee guida miravano a orientare lo sviluppo di un'IA che fosse eticamente responsabile e socialmente accettabile. È da sottolineare come nel terzo capitolo venga fornita un *Assessment List* che mira ad agevolare la verifica dell'aderenza del sistema oggetto di valutazione ai sette requisiti descritti nelle linee guida.

Una tappa fondamentale dell'iter legislativo si ha nell'aprile del 2021 quando viene presentata dalla Commissione europea una prima proposta di Regolamento sull'Intelligenza Artificiale, nota come AI ACT. Il percorso si presenta fin dal principio particolarmente complesso poiché le Istituzioni si sono scontrate con la difficoltà a far collimare le esigenze dei vari Stati europei sulle questioni critiche che attengono i temi di governance dell'IA, al contempo fronteggiando le pressioni provenienti dalle diverse lobby presenti sul territorio europeo. A un anno e mezzo dalla proposta fatta dalla Commissione Europea, il 6 dicembre del 2022, il Consiglio Europeo rende il proprio orienta-

mento generale riguardo il Regolamento per assicurarsi che i sistemi di IA introdotti nel mercato dell'UE e utilizzati all'interno dell'Unione siano sicuri e conformi alla legislazione esistente su diritti fondamentali e valori europei. Il successo dell'adozione dell'orientamento generale del Consiglio Europeo ha portato alla fase successiva dell'iter legislativo europeo, i "triloghi". Essi rappresentano una fase di discussione non pubblica ed hanno la funzione di favorire l'accordo tra Consiglio Europeo, Commissione e Parlamento. Normalmente, durante questa fase, il testo della normativa da approvare viene esaminato punto per punto per giungere ad un risultato che soddisfi le esigenze e gli interessi delle varie parti coinvolte. Nel caso dell'IA ACT, la fase dei "triloghi" si è rivelata particolarmente irta di ostacoli, soprattutto di natura politica, e ha condotto a radicali modifiche al testo rispetto alla formulazione originaria. Solo dopo lunghe trattative, infatti, il 9 dicembre 2023 si è giunti ad un'intesa politica tra Consiglio e Parlamento europeo sulla proposta. L'intesa si è fondata nell'aver trovato un punto di equilibrio su alcuni aspetti critici del testo: individuazione di regole per i modelli di IA per finalità generali ad alto impatto che possono comportare rischi sistemici in futuro e per i sistemi di IA considerati ad alto rischio; revisione del sistema di governance con alcuni poteri di esecuzione a livello dell'UE; ampliamento dell'elenco dei divieti, ma al contempo apertura alla possibilità di impiegare, con alcuni limiti, l'identificazione biometrica remota da parte delle autorità di contrasto negli spazi pubblici; miglioramento della protezione dei di-

ritti tramite l'obbligo per gli operatori di sistemi di IA ad alto rischio di effettuare una valutazione d'impatto sui diritti fondamentali.

Con l'approvazione dell'accordo politico da parte del COREPER (Comitato dei rappresentanti permanenti dell'UE), il 2 febbraio del 2024, gli Stati membri hanno approvato la proposta di testo dell'AI ACT concordata dal trologo del 9 dicembre 2023.

Uno degli ultimi tasselli del quadro è stato aggiunto il 13 marzo 2024, in cui il Parlamento Europeo in seduta plenaria, anticipando la votazione calendarizzata per fine aprile, ha approvato il Regolamento con 523 voti favorevoli, 46 contrari e 49 astensioni.

Il testo approvato dal Parlamento, così come spiegato nel Comunicato stampa del 13/3/24, è stato sottoposto alla verifica finale da parte dei giuristi-linguisti. Il 21/5/2024 arriva poi la definitiva approvazione formale da parte del Consiglio. Il regolamento dovrà ora essere pubblicato nella Gazzetta ufficiale dell'UE e si applicherà 24 mesi dopo l'entrata in vigore, con alcune eccezioni (i divieti relativi a pratiche proibite si applicheranno già a partire da sei mesi dopo l'entrata in vigore, i codici condotta dovrebbero essere pubblicati entro nove mesi dall'entrata in vigore, le norme sui sistemi di IA per finalità generali, compresa la governance 12 mesi dopo, infine gli obblighi per i sistemi ad alto rischio 36 mesi dopo).

## L'APPROCCIO BASATO SUL RISCHIO AL CENTRO DEL REGOLAMENTO EUROPEO SULL'IA

Il filo conduttore del Regolamento è rappresentato dall'approccio basato sul rischio (o *risk-based approach*), che affonda le proprie radici giuridiche nel principio di precauzione. Questo principio permette di prendere decisioni in situazioni incerte, gestendo rischi e pericoli potenziali ancora indefiniti, specialmente rilevante per l'intelligenza artificiale, ambito ancora avvolto da incertezza scientifica con rischi e impatti difficili da prevedere.

La normativa divide i sistemi di IA in livelli di rischio inaccettabile, alto e limitato. Il livello del rischio autorizza o meno la vendita, messa in funzione e uso sul territorio dell'UE del dispositivo di IA. Quando un sistema di IA presenta un rischio inaccettabile, esso è proibito sul territorio dell'Unione e non potrà circolare nel mercato europeo. Rientrano tra i **sistemi proibiti** quelli che, per distorcere il comportamento delle persone: i) usano tecniche manipolatorie o ingannevoli o ii) sfruttano le loro vulnerabilità; iii) sistemi di social scoring che associno un punteggio sociale a determinate caratteristiche personali o comportamenti; iv) sistemi che valutano il rischio di commissione di reati basandosi esclusivamente su attività di profiling o su analisi della personalità; v) sistemi che raccolgono dati di riconoscimento facciale da internet o da CCTV (cosiddetto scraping); vi) i sistemi di riconoscimento emotivo usati in contesti di lavoro o educativi; vii) i sistemi di categorizzazione biometrica per inferire categorie speciali di dati personali (come orientamento sessua-

le, opinioni politiche, credenze religiose o filosofiche); viii) il riconoscimento facciale in tempo reale in luoghi di pubblico accesso per finalità di applicazione della legge (salvo le eccezioni previste dalla legge);

Per gli altri sistemi di IA, la normativa introduce vari livelli di compliance a seconda della gravità del rischio.

Per i sistemi di General Purpose AI (GPAI), si introducono precisi obblighi di trasparenza, mantenimento di documentazione tecnica aggiornata, e di stesura di policy per il rispetto del diritto d'autore europeo. Inoltre, se i GPAI sono classificati come portatori di un rischio sistemico per le persone e la società, i *provider* devono valutare il modello, valutare e mitigare i possibili rischi sistemici, tenere traccia di incidenti gravi e delle misure correttive attuate, nonché assicurare un adeguato livello di cybersecurity.

Per i sistemi di IA ad alto rischio (elencati nell'Annex III e tra cui rientrano, ad esempio, i sistemi di IA usati come componenti di sicurezza nella gestione di infrastrutture digitali critiche, oppure per l'assunzione del personale), si prevede che sia implementato, documentato e mantenuto un sistema di gestione del rischio, che vengano seguite pratiche di data governance appropriate per gli usi previsti, che sia mantenuta aggiornata documentazione tecnica inerente il sistema, che sia assicurata sufficiente trasparenza circa la sua operatività, che sia reso possibile sorvegliare efficacemente il sistema da parte di operatori umani, e che sia garantito un livello appropriato di accuratezza, robustezza e sicurezza informatica per tutto il ciclo di vita.

Altri requisiti devono essere soddisfatti (tra cui: la necessità di sottoporre il sistema a una procedura di conformità) a seconda che si parli di *provider* (secondo l'art. 3, persona fisica o giuridica, un'autorità pubblica, un'agenzia o un altro organismo che sviluppa un sistema di IA o un modello di IA per scopi generali o che fa sviluppare un sistema di IA o un modello di IA per scopi generali e li immette sul mercato o mette in servizio il sistema con il proprio nome o marchio, a titolo oneroso o gratuito) o *deployer* (secondo l'art. 3, “persona fisica o giuridica, autorità pubblica, agenzia o altro organismo che utilizza un sistema di IA sotto la propria autorità, tranne nel caso in cui il sistema di IA sia utilizzato nel corso di un'attività personale non professionale”) di IA ad alto rischio e si rimanda al testo per i dettagli. Ai fini di questo lavoro, si intende portare l'attenzione sull'obbligo di valutazione dell'impatto sui diritti fondamentali (o *fundamental rights impact assessment* – FRIA) imposto a determinati *deployer* (ossia, secondo l'art. 27, organismi di diritto pubblico, soggetti privati che forniscono servizi pubblici, nonché *deployer* di sistemi ad alto rischio che siano destinati a valutare la solvibilità delle persone fisiche e/o a determinare i prezzi delle assicurazioni ad esse associate) che facciano uso di alcuni sistemi di IA ad alto rischio elencati nell'Allegato III, prima di procedere all'immissione sul mercato. Questo obbligo, alleggerito nel suo contenuto rispetto alla versione precedente del testo, è il risultato del compromesso raggiunto nei tri loghi, e confermato dal Parlamento il 13 marzo 2024 e dal Consiglio il 21 maggio 2024.

La valutazione permette a chi implementa un sistema di IA ad alto rischio, dato lo scopo previsto del sistema, di definire e di identificare le categorie di persone che potrebbero essere influenzate dall'uso del sistema. L'obiettivo della valutazione è riconoscere i rischi specifici di danni e impatti per le categorie determinate e, di conseguenza, stabilire le misure appropriate per mitigare tali rischi. La mitigazione potrà riguardare anche la predisposizione di misure di governance interna, la quale assume dunque un'importanza chiave nell'adozione responsabile delle tecnologie di IA.

La valutazione dell'impatto prevista dall'AI ACT è un passo decisivo verso lo sviluppo di sistemi di IA ad alto rischio affidabili nel quadro europeo di principi e di diritti. Non solo: si presta a essere una via maestra per chiunque intenda progettare, sviluppare e implementare IA, sia per processi interni, sia per il miglioramento della propria offerta commerciale verso clienti business e consumer, a prescindere dal livello di rischio. La valutazione d'impatto sui diritti fondamentali fornisce infatti sia una mappa del territorio per chi conduce l'analisi, aiutando a riconoscere i punti di criticità in anticipo e a intervenire tempestivamente, sia una risorsa verso l'esterno, alimentando la fiducia dei clienti e della società, i quali potranno avvalersi di un sistema, prodotto o servizio che tratta l'IA e le delicate tematiche connesse (tra cui: gestione dei dati personali, mitigazione del rischio di discriminazioni, impatti ambientali) con un alto grado di responsabilità.

## PANORAMICA DELLE VALUTAZIONI D'IMPATTO

Partendo da queste premesse, in quanto segue esporremo una panoramica degli strumenti di valutazione dell'impatto algoritmico (*Algorithmic Impact Assessment*, AIA) creati e messi sinora a disposizione da alcune istituzioni, sia governative sia private. Il resoconto che presentiamo non ha la pretesa di essere esaustivo e non è escluso che nuove e ulteriori valutazioni dell'impatto siano in preparazione e/o usate al momento della redazione di questo documento.

Nelle sezioni successive, forniremo una definizione dell'AIA e descriveremo i vari approcci adottati dai modelli di valutazione dell'impatto, evidenziando le diverse fasi di cui si compongono.

Mentre attendiamo il rilascio del modello di questionario dall'Ufficio Europeo dell'IA, un ente creato in base all'AI ACT con il compito di facilitare l'adempimento dei requisiti di valutazione d'impatto per alcuni sistemi di intelligenza artificiale ad alto rischio, questo documento fornisce una panoramica utile. Il modello include un'analisi dell'impatto sui diritti fondamentali e assicura la sua corretta applicazione. Questa panoramica è pensata per supportare chi intende sviluppare, distribuire o adottare sistemi di IA, mantenendo alta l'attenzione sulle relative questioni etiche e normative. La prospettiva è di interesse anche per le organizzazioni che prevedono di intraprendere percorsi di certificazione, come quello previsto dalla recente norma ISO 42001:23 relativa al sistema di gestione dell'IA, dove la valutazione d'impatto è parte integrante dei requisiti da soddisfare.



## COSA SONO LE VALUTAZIONI DELL'IMPATTO ALGORITMICO?

Le valutazioni dell'impatto algoritmico sono strumenti utili per l'individuazione dei rischi che possono emergere con l'utilizzo dei sistemi di IA e per l'adozione di misure ragionevoli a mitigare i rischi individuati. Questi strumenti vengono denominati in diversi modi, tra cui "valutazione dell'impatto dell'Intelligenza Artificiale", "valutazione dell'impatto dei sistemi decisionali automatizzati" o "valutazione dell'impatto algoritmico". Nel documento, si è scelto di usare "valutazione dell'impatto algoritmico" (o l'acronimo AIA, dall'inglese Algorithmic Impact Assessment) per riferirsi a queste valutazioni.

Gli AIA funzionano in modo simile alla valutazione dell'impatto della protezione dei dati (*Data Protection Impact Assessment* – DPIA), una procedura che mira a descrivere un trattamento di dati personali per valutarne la necessità e la proporzionalità nonché i relativi rischi, con lo scopo di mettere in atto misure idonee ad affrontarli, richiesta all'articolo 35 del Regolamento UE n. 2016/679 in materia di protezione dei dati personali (anche noto come GDPR, da *General Data Protection Regulation*). In particolare, l'AIA è utile per aiutare sviluppatori, distributori e coloro che desiderano utilizzare sistemi basati sull'Intelligenza Artificiale ad essere consapevoli dei rischi che possono comportare questi sistemi. Consentono, inoltre, di valutare l'impatto del sistema sugli aspetti etici e legali con lo scopo di mitigare potenziali rischi.

La comprensione dei rischi e l'adozione di misure ragionevoli e proporzionate dovrebbero garantire che

questi sistemi siano basati su uno *human centric approach*, riducendo gli impatti negativi e massimizzando i benefici per le persone e la società nel suo insieme. Generalmente è consigliato effettuare la valutazione dell'impatto sia durante lo sviluppo del sistema sia prima della sua distribuzione (per prevenire in tempo errori che potrebbero essere, in seguito, costosi), ma anche periodicamente per valutare le conseguenze del sistema di Intelligenza Artificiale nel tempo.

Come detto in precedenza, anche l'AI ACT nella sua versione attuale prescrive per alcuni *deployer* di sistemi di IA rientranti tra quelli classificati come ad alto rischio (per es. nel settore bancario e assicurativo) l'obbligo di effettuare una valutazione dell'impatto sui diritti fondamentali (*fundamental rights impact assessment*, FRIA).

## PRINCIPALI FASI DI AIA

Anche se i modelli di AIA variano in termini di approccio, focus e completezza, possiamo osservare che si conformano generalmente a un processo in cinque fasi: descrizione del sistema, valutazione del rischio, valutazione dell'impatto, identificazione delle misure di mitigazione e metodi per la revisione periodica del sistema.



**Descrizione del sistema.** In questa fase viene generalmente chiesto di descrivere: il progetto (scopo, usi previsti, benefici dell'utilizzo di un sistema decisionale automatizzato); il modello (caratteristiche del sistema, input e output, il livello di autonomia del sistema); i dati (origine del set di dati sia in uso sia quelli di addestramento); gli stakeholders coinvolti.

**Valutazione del rischio.** Questa fase generalmente differisce nel metodo: alcuni AIA chiedono semplicemente di interrogarsi sui danni che può causare l'utilizzo del sistema algoritmico, altri invece si basano su metriche di valutazione del rischio. In quest'ultimo caso può essere chiesto di rispondere a domande a scelta multipla e dall'incrocio delle risposte viene rilevato in automatico il rischio oppure viene presentato un elenco di caratteristiche dei sistemi ad alto rischio e viene chiesto di indicare se il proprio sistema le rispecchia (quest'ultimo metodo di valutazione del rischio è simile all'approccio di definizione del rischio della Commissione Europea indicata nella Proposta di regolamentazione dell'IA).

**Valutazione dell'impatto.** Questa fase è dedicata a valutare l'impatto umano, sociale e ambientale tramite l'analisi sull'influenza sull'autonomia umana, sulla dignità umana, sul diritto alla privacy, sull'equità e sulla non discriminazione, sulla coesione sociale (per valutare queste dimensioni può essere utile lo "*Stakeholders impact assessment*" dell'Alan Turing Institute [Leslie 2019]) e se il sistema influirà sulla sostenibilità ambientale (valutando ad esempio l'energia necessaria per far funzionare il sistema o possibili impatti futuri

sull'ambiente). In questa fase viene anche richiesto di riflettere sulla sicurezza del sistema e sulla sua robustezza. Viene inoltre richiesto di valutare la trasparenza (quindi apertura su dati, algoritmi e processo) del sistema e l'accountability.

**Identificazione delle misure di mitigazione.** Lo scopo delle AIA non dovrebbe essere soltanto quello di valutare rischi e impatti dei sistemi algoritmici, bensì quello di identificare adeguate misure di mitigazione. Non tutte le AIA presenti attualmente in letteratura sono state create per questo scopo.

**Revisione periodica del sistema.** Considerando che alcuni impatti possono essere identificati solamente dopo l'implementazione del sistema, viene richiesta la valutazione periodica al fine di poter valutare le conseguenze del sistema nel tempo. Una volta identificati nuovi impatti, inoltre, è necessario identificare e mettere in atto appropriate misure di mitigazione.

PARTE II

GLI STRUMENTI



## INTRODUZIONE

Come presentato nella parte I, eseguire la valutazione d'impatto dei rischi dei sistemi di IA è, e sarà sempre più, un passaggio fondamentale per le organizzazioni che vogliono agire in maniera responsabile, trasparente e socialmente sostenibile. Inoltre, sulla base dell'indirizzo legislativo indicato dall'ultima versione dell'AI ACT, in certi casi l'esecuzione di una valutazione dell'impatto sui diritti fondamentali non sarà più una scelta opzionale delle organizzazioni: per alcune tipologie di deployer e domini di applicazione di IA ad alto rischio, sarà un obbligo.

È in questo scenario che istituzioni governative, organizzazioni internazionali ed europee, società private e mondo accademico hanno avviato da alcuni anni lo studio e la sperimentazione sui modelli di valutazione degli impatti del sistema di IA: accomunati dall'approccio basato sul rischio, questi modelli sono stati declinati in modi diversi a seconda dei contesti e del retroterra degli stessi promotori.

I risultati, presentati attraverso pubblicazioni accademiche e non, sono stati la definizione di percorsi di valutazione che, seppur non omogenei per approccio, focus e completezza, puntano a essere, per chi li usa, strumenti adatti a valutare potenziali rischi e impatti dei sistemi di IA sull'organizzazione, sugli individui e categorie di individui, sulla società nel suo complesso, durante tutto il ciclo di vita del sistema.

Di seguito, verranno proposte le descrizioni di alto

livello di un elenco, non esaustivo, di modelli di valutazione d'impatto dei rischi dei sistemi di IA creati o adottati da diversi attori, pubblici e privati. Offrendo una panoramica delle caratteristiche principali di alcuni degli strumenti esistenti nel momento in cui questo documento viene redatto, si vuole:

- mostrare come possano essere **misurati, con approcci diversi, gli impatti dell'IA** su individui, organizzazioni e società;
- **sottolineare l'importanza di strumenti che operano sul piano della consapevolezza, prevenzione e mitigazione di potenziali rischi e impatti**, documentandone l'esistenza stessa e la cura profusa per la loro implementazione e distribuzione (per lo più senza scopo di lucro);
- **spingere i decisori aziendali** a domandarsi non solo se è il caso di adottare sin da subito un approccio basato sulla mitigazione dei rischi ma anche **a scegliere il modello più valido** o, addirittura, crearne uno che possa recepire e rielaborare le best practice degli strumenti esistenti.

L'elenco si presta ad essere aggiornato periodicamente, in attesa, per i paesi dell'UE, dello strumento che verrà pubblicato dall'AI Office europeo.



## CANADIAN ALGORITHMIC IMPACT ASSESSMENT (AIA) TOOL (GOVERNO DEL CANADA)

Il **Canadian Algorithmic Impact Assessment** [1] è uno strumento di valutazione dell'impatto reso **obbligatorio** dal Governo del Canada **per la pubblica amministrazione** ai sensi della “**Directive on Automated Decision-Making**” entrata in vigore in data 01.04.2019.

Lo strumento è costituito da un **questionario** che determina il livello di impatto di un sistema decisionale automatizzato. La sua utilità emerge principalmente nella fase di progettazione di soluzioni che fanno uso di IA vale a dire quando si vogliono prevenire “by design” gli impatti negativi sugli utilizzatori dei sistemi o su chi eventualmente ne subisce le decisioni.

Il questionario si compone di **51 domande sul rischio** e **34 sulla mitigazione**. I punteggi di valutazione si basano su molti fattori, tra cui la progettazione del sistema, l'algoritmo, il tipo di decisione, l'impatto e i dati.

I punteggi di valutazione riguardano sei **aree di rischio** (il progetto, il sistema e le sue caratteristiche, l'algoritmo, il tipo di decisione, l'impatto, i dati) e due **aree di mitigazione** (consultazioni con stakeholder interni e esterni; misure di de-risking e mitigazione, tra cui la qualità dei dati perché siano rappresentativi e imparziali, le procedure in atto per controllare il sistema e le sue decisioni, le misure prese per salvaguardare le informazioni personali).

In base alle risposte date, viene **in automatico attribuito un punteggio** su scala percentuale e **restituito il risultato del livello di impatto** del sistema:

- Livello I – Impatto minimo o nullo;
- Livello II – Impatto moderato;
- Livello III – Impatto elevato;
- Livello IV – Impatto molto elevato.

Il risultato include indicazioni sulle aree che necessitano di una revisione e suggerimenti sulle misure di mitigazione del rischio da adottare.

Secondo la Direttiva, i risultati devono essere pubblicati e resi accessibili.

A dimostrazione della rilevanza dello strumento e della necessità che rimanga al passo non solo con l'innovazione tecnologica ma anche con quella socio-culturale, il questionario ha subito alcuni importanti aggiornamenti negli anni che hanno posto l'accento su diritti e libertà fondamentali, uguaglianza, dignità e autonomia degli individui, senza trascurare le questioni legate alla salute e all'ambiente.

## ALGORITHMIC IMPACT ASSESSMENT (AIA) (FEDERAL CIO COUNCIL)

Nel 2021 il *Federal Chief Information Officer* (CIO) *Council* degli Stati Uniti (ovvero il Consiglio degli Stati Uniti che mira a migliorare le pratiche relative alla progettazione, acquisizione, sviluppo, modernizzazione, uso e condivisione delle risorse informative federali), pubblica sul suo sito uno **strumento di AIA** in versione Alpha [2].

Lo strumento è stato sviluppato per aiutare le agenzie del governo federale statunitense a valutare i rischi associati all'utilizzo di sistemi decisionali automatizzati. Essendo una versione Alpha, chi utilizza lo strumento può fornire *feedback* e pareri sulla piattaforma Github. L'efficacia di questo strumento si manifesta soprattutto nelle **fasi iniziali di un progetto**, offrendo un supporto fondamentale per valutarne impatti e rischi sin dall'avvio.

Lo strumento è costituito da un **questionario** composto da **dodici sezioni**. Le prime otto sezioni pongono domande relative a: obiettivo del progetto, autonomia del sistema, dettagli del fornitore, privacy e sicurezza dei dati, tipo di dati e origine degli stessi, qualità dei dati, informazioni sull'algoritmo, valutazione del sistema e profilo di rischio. Le ultime quattro sezioni pongono domande riguardanti le misure di mitigazione e riduzione del rischio messe in atto (in particolare, le domande sono relative alla qualità dei dati, alla privacy, al coinvolgimento degli stakeholders, all'*accountability* e tracciabilità del sistema).

Questa valutazione è composta **in prevalenza da**

**domande a risposta libera.** Le domande a risposta dicotomica servono principalmente da sbarramento (in base alla risposta data, infatti, possono comparire ulteriori domande a risposta aperta). Per **valutare il rischio** del sistema vengono proposte alcune domande a scelta multipla.

In termini metodologici, lo strumento è principalmente di tipo qualitativo, fondandosi su domande aperte per stimolare la riflessione sull'impatto del sistema, ma presenta anche domande chiuse per rilevare il rischio (metodo quantitativo).

Essendo in versione Alpha, attualmente non viene fornito in output il livello di impatto del sistema né suggerite eventuali misure di mitigazione.

## AI SYSTEMS ETHICS SELF-ASSESSMENT (DIGITAL DUBAI AUTHORITY)

La **Digital Dubai Authority** è un ente fondato dallo sceicco Mohammed Bin Rashid Al Maktoum con lo scopo di sviluppare e supervisionare l'attuazione di politiche e strategie che regolino le questioni relative a dati, trasformazione digitale, sicurezza informatica e tecnologie della città di Dubai. Nel 2020, la Digital Dubai Authority ha reso disponibile sul suo sito uno **strumento di autovalutazione dell'eticità dell'IA** [3].

Basato sulle linee etiche dell'IA di Dubai, lo strumento è stato sviluppato per aiutare le organizzazioni di sviluppatori o fornitori a **valutare il livello etico di un sistema di IA**. Anche per questo strumento si può affermare che è di grande valore per analizzare rischi e impatti nelle fasi preliminari di un progetto.

Unisce metodo quantitativo e qualitativo in quanto presenta sia domande chiuse per rilevare il rischio sia domande aperte che stimola la riflessione sull'impatto del sistema.

Il sistema di autovalutazione consiste in un questionario che, in base alle risposte date, determina il livello di impatto complessivo del sistema e il dettaglio sul livello di impatto per ogni dimensione considerata (*fairness, accountability, transparency ed explainability*).

All'inizio del questionario, è richiesto di descrivere il sistema di IA e le sue finalità e di specificare se esso si occupa di decisioni non significative, significative o critiche, andando così ad autovalutare il rischio del sistema.

Successivamente vengono poste domande relative a:

1) equità (*fairness*); 2) accountability; 3) trasparenza; 4) spiegabilità (*explainability*). Il numero e la tipologia di domande proposte dipendono dal risultato della classificazione del sistema (ad esempio se il sistema è stato classificato ad alto rischio, verranno proposte tutte le domande del questionario; se il sistema è stato classificato a basso rischio lo strumento proporrà solo le domande rilevanti per quel tipo di rischio).

Alla fine del questionario viene espresso in percentuale il livello di impatto complessivo del sistema e il dettaglio dell'impatto per ogni principio etico (equità, accountability, trasparenza, spiegabilità).

Nelle parti che necessitano di ulteriori revisioni viene consigliato di consultare le linee guida e le relative misure di mitigazione da mettere in atto al fine di rendere etico il sistema.

FROM PRINCIPLES TO PRACTICE: AN INTERDISCIPLINARY  
FRAMEWORK TO OPERATIONALIZE AI ETHICS  
(AI ETHICS IMPACT GROUP)

**AI Ethics Impact Group (AIEI Group)** è un consorzio interdisciplinare, con sede in Germania, guidato dalla VDE Association for Electrical, Electronic & Information Technologies e dalla Bertelsmann Stiftung (fondazione indipendente di diritto privato). AIEI Group si pone l'obiettivo di assicurarsi che l'etica dell'IA sia facile da comprendere per tutte le organizzazioni e i cittadini d'Europa. A tal proposito, nel 2020, propone un **framework** [4] specificamente progettato **per tradurre i principi etici in azioni pratiche** da affrontare durante la progettazione, l'implementazione e la valutazione dei sistemi di IA.

Il framework **proposto segue il modello VCIO (values, criteria, indicators, and observables)**. Vengono quindi definiti alcuni valori specifici e desiderabili affinché un sistema di AI sia considerato etico (trasparenza, accountability, privacy, equità, affidabilità e sostenibilità). I valori vengono scomposti in criteri (ad esempio il valore "transparency" viene scomposto in apertura sull'origine del set di dati e sull'algoritmo o modello utilizzato). Ogni criterio viene a sua volta distinto in indicatori, ovvero una serie di domande a scelta multipla utili per poter valutare il criterio stesso.

Per ogni risposta scelta viene assegnata una lettera su una scala di valutazione da A a G (dove A indica il miglior soddisfacimento possibile del criterio). Per rendere il risultato della valutazione del sistema di IA, viene utilizzata una "**etichettatura etica**", simile all'e-

tichetta di classificazione energetica. Per ogni criterio, quindi, sarà possibile osservare la classificazione etica dell'IA e capire quali dimensioni del sistema necessitano di ulteriori revisioni.

Si conforma alle **Ethics Guidelines for Trustworthy Artificial Intelligence** pubblicate nel 2019 dall'High-Level Expert Group on AI della Commissione Europea, ma non è di utilità per chi volesse anche identificare misure di mitigazione.



## ARTIFICIAL INTELLIGENCE IMPACT ASSESSMENT (CONSIDERATI / ECP)

**Considerati** (società di consulenza legale con lo scopo di aiutare le organizzazioni ad innovarsi con l'uso responsabile di tecnologia e dati con sede in Olanda) e la **Platform voor de InformatieSamenleving** (ECP, piattaforma indipendente e neutrale in cui collaborano sia il settore pubblico sia il settore privato con lo scopo di scambiarsi conoscenze sulla progettazione responsabile di una società digitalizzata) nel 2019 propongono e pubblicano una **valutazione dell'impatto del sistema di IA** [5] composta da otto step.

Il primo step consta di una serie di **domande di screening** alle quali si può rispondere su scala dicotomica sì/no. Se si risponde affermativamente anche solo a una di queste domande, si consiglia procedere nella valutazione dell'impatto perché il sistema rientra tra quelli considerati ad alto rischio.

Gli step intermedi, dal 2 al 5, prevedono di rispondere **domande aperte**: sul progetto, sui benefici del sistema, sull'impatto della tecnologia su valori umani e sociali (si riferisce ai principi di dignità umana, autonomia umana, responsabilità, trasparenza, equità, democrazia, sicurezza, privacy e protezione dei dati, sostenibilità), sulla sicurezza, trasparenza e affidabilità.

Per giungere alla valutazione di eticità dell'applicazione, lo step 6 richiede di **considerare congiuntamente benefici e rischi individuati** negli step precedenti. Gli step finali, 7 e 8, sono rispettivamente dedicati alla documentazione della valutazione e alla rivalutazione periodica del sistema.

Per ogni step, oltre alle domande a cui si deve rispondere, è presente una descrizione di come dovrebbe essere un sistema di IA per essere considerato affidabile e quindi per funzionare correttamente.

Questo strumento di valutazione si conforma alle **Ethics Guidelines for Trustworthy Artificial Intelligence** dell'AI HLEG.

## UNLOCKING PUBLIC SECTOR AI. AI PROCUREMENT IN A BOX (WORLD ECONOMIC FORUM)

Il **World Economic Forum** (WEF) è una fondazione internazionale indipendente, con sede in Svizzera, che si propone di abilitare la cooperazione tra pubblico e privato per facilitare il progresso comune di fronte alle sfide globali. Nel contesto dell'IA, il WEF intende guidare le iniziative di co-progettazione, test e implementazione di politiche che facciano leva sui potenziali delle tecnologie di IA/ML, affrontandone al contempo le problematiche.

Nel **2020** propone e pubblica uno **strumento di valutazione dei potenziali rischi connessi all'adozione di qualsiasi soluzione che integra componenti di IA** [6], inserito in un più ampio programma per liberare il potenziale dell'IA nel settore pubblico (*Unlocking Public Sector AI. AI Procurement in a Box*) con specifiche Linee guida.

La valutazione è composta da una **lista di controllo** dedicata alla valutazione del rischio. Quando si risponde "sì" ad una domanda, vengono proposte ulteriori domande (a risposta aperta) nella sezione successiva per determinare l'impatto della tecnologia.

Oltre a rispondere alle domande richieste nelle due liste di controllo, viene richiesto di **descrivere il sistema** (quindi lo scopo, le funzioni, i dati utilizzati, l'architettura del sistema, stakeholders coinvolti, benefici dell'utilizzo del sistema).

Il documento è pensato per utenti non esperti: mira a **supportare la riflessione all'interno del settore pubblico prima di procedere all'acquisto di IA da fornito-**

**ri terzi** così da assicurare l'approvvigionamento (*procurement*) di sistemi affidabili.

Nei due anni successivi alla pubblicazione dello strumento sono stati avviati progetti pilota in diversi stati. Il Brasile, ad esempio, ha applicato questa metodologia per supportare il processo di procurement di tecnologie di IA dell'azienda sanitaria University of São Paulo's Hospital das Clínicas (HC) e per l'azienda del trasporto Metrô de São Paulo (WEF 2022).

## AUTOMATED DECISION IMPACT ASSESSMENT (ADIA) FRAMEWORK (OPEN LOOP/FACEBOOK)

**Open Loop** è un'iniziativa privata che si basa sul lavoro di Facebook/Meta per aiutare a sviluppare raccomandazioni politiche sull'uso della tecnologia. L'iniziativa si propone di creare e testare prototipi di leggi e regolamenti prima che essi vengano emanati, in modo tale che gli stakeholder possano comprendere il loro funzionamento nel mondo reale.

Nel **2020** viene pubblicato un documento [7] che propone per il contesto europeo un **prototipo di Automated Decision Impact Assessment (ADIA) framework** con l'obiettivo di testare sia la comprensione di un prototipo di legge (che richiede una valutazione del rischio e dell'impatto dei sistemi di IA) sia il procedimento di una valutazione dell'impatto.

L'obiettivo dell'esperimento è quello di aiutare sviluppatori e utenti di Intelligenza Artificiale ad essere consapevoli dei rischi dei sistemi di IA. Scopo dell'esperimento è anche quello di aiutare ad individuare adeguate modalità di mitigazione dei potenziali rischi causati dal sistema.

La valutazione dell'impatto è composta da dieci passaggi:

1. Descrivere il sistema.
2. Valutare come il sistema cambierà la situazione esistente, stimolando la riflessione sui vantaggi e sui rischi che può comportare l'adozione del sistema.
3. Considerare gli aspetti negativi (ad esempio opacità del sistema).
4. Pensare all'impatto sugli stakeholders (con **tabella**

- di tassonomia dei danni** a supporto).
5. Valutare le tensioni che possono nascere tra i benefici dell'azienda sviluppatrice/produttrice (ad esempio, profitti maggiori) e i rischi per gli stakeholders (ad esempio, l'algoritmo spinge gli utenti a comportamenti non necessariamente favorevoli per loro, come fare microtransazioni sulla piattaforma). Oltre alla tabella di tassonomia dei danni, è fornita **tabella di panoramica dei valori e principi** per supportare la riflessione.
  6. Determinare la probabilità che si verifichi realmente un impatto negativo.
  7. Identificare misure di mitigazione del danno (con tabella che **elenca misure di mitigazione sia tecniche sia organizzative** a supporto).
  8. Valutare le conseguenze dopo l'applicazione delle misure di mitigazione (o dopo altre eventuali modifiche).
  9. Decidere quali modifiche o misure di mitigazione verranno implementate.
  10. Implementazione e documentazione.

Nell'esperimento **sono state coinvolte diverse aziende** che adottano o sviluppano IA per consentire meccanismi di feedback e miglioramento continuo.

Rispetto ad altri strumenti che hanno la forma di check-list più o meno dettagliate, questo modello ha un approccio procedurale e più completo alla valutazione degli impatti. Anche questo strumento si ispira e conforma alle *Ethics Guidelines for Trustworthy Artificial Intelligence*, pubblicate nel 2019 AI HLEG.

## AUTOMATED DECISION MAKING SYSTEMS IN THE PUBLIC SECTOR. AN IMPACT ASSESSMENT TOOL FOR PUBLIC AUTHORITIES (ALGORITHM WATCH)

**Algorithm Watch** è un'organizzazione di ricerca senza scopo di lucro con sede in Germania che si pone l'obiettivo di valutare e far luce sui processi decisionali algoritmici che hanno una rilevanza sociale, ossia che vengono utilizzati per prevedere o prescrivere l'azione umana o per prendere decisioni automaticamente. Nel **2021**, l'organizzazione sviluppa e pubblica uno **strumento di valutazione dell'impatto dei sistemi decisionali automatizzati implementati nel settore pubblico**, basato sui **sette requisiti UE** [8].

La procedura di valutazione proposta consiste in uno strumento pratico per generare trasparenza ed è composta dalle due fasi riportate qui di seguito:

- Nella **prima fase** viene presentata una **lista di controllo** ("**transparency triage**") che aiuta a determinare le principali questioni di trasparenza inerenti al progetto che devono essere documentate. Si raccomanda di compilare questa lista di controllo **durante la fase di pianificazione** per prendere in considerazione ulteriori specifiche per il progetto, oltre ai suoi obiettivi primari. Dalle risposte fornite alle domande della lista di controllo può emergere che il sistema necessita di ulteriori requisiti di trasparenza, da trattare nel dettaglio nella fase di valutazione successiva rispondendo a domande specifiche.
- Nella **seconda fase** viene chiesto di redigere una relazione sulla trasparenza. Per aiutare nella redazione della relazione viene proposta una **seconda**

**lista di controllo.**

Sebbene questo strumento non includa né un sistema per valutare l'impatto né una definizione del rischio e non aiuti a identificare possibili misure di mitigazione, le due liste di controllo di cui è composto contribuiscono a **generare trasparenza** su progetti e applicazioni di **automazione tecnologica** nella **pubblica amministrazione**. La **relazione conclusiva sulla trasparenza** serve infatti a dimostrare che le questioni etiche più importanti sono state riconosciute e portate sotto controllo umano, e che è stata garantita un'adeguata responsabilità del processo.



## ALGORITHMIC IMPACT ASSESSMENT (ADA LOVELACE INSTITUTE)

L'**Ada Lovelace Institute** è un istituto di ricerca indipendente con sede nel Regno Unito che ha l'obiettivo di garantire che la tecnologia sia al servizio dell'umanità e che dati e IA siano utilizzati per massimizzare il benessere sociale. Nel **2022**, l'istituto ha proposto e pubblicato uno strumento di valutazione dell'impatto algoritmico (AIA) specifico per i sistemi di IA applicati nel **settore sanitario** [9].

Questo AIA è oggetto di un **progetto pilota** che coinvolge il **Sistema Sanitario Nazionale del Regno Unito** (National Health Service, NHS). Le applicazioni di IA che vorranno accedere ai dati della National Medical Imaging Platform (NMIP) del NHS dovranno infatti prima essere sottoposte alla valutazione dell'impatto algoritmico. Un Comitato esaminerà la valutazione e deciderà se autorizzare o meno l'accesso alla NMIP.

La valutazione dell'impatto algoritmico prevede tre esercizi:

1. Un **esercizio di riflessione** iniziale sui potenziali impatti del progetto (svolto dal *team* di progetto). Viene fornito in questa fase un documento nel quale sono proposte una serie di domande alle quali rispondere. Ad esempio viene richiesto: di descrivere il progetto, il modello, gli stakeholders; di riflettere su possibili disuguaglianze e discriminazioni; di considerare l'impatto ambientale e la possibilità di ricorso; di riflettere sull'autonomia umana; di interrogarsi sulle sfide e sugli ostacoli che potrebbero verificarsi; di riflettere sullo scenario migliore e sul-

lo scenario peggiore che potrebbero verificarsi; di individuare i potenziali danni dopo l'implementazione e di individuare le relative misure di mitigazione.

2. La **partecipazione ad un workshop** dove il *team* del progetto può ricevere *feedback* (da un gruppo di pazienti e operatori sanitari) sul documento generato nello step iniziale.
3. Una **sintesi finale** dei due step precedenti che verrà poi presentata al Comitato che deciderà se autorizzare l'accesso alla NMIP.

CAP AI- A PROCEDURE FOR CONDUCTING CONFORMITY ASSESSMENT OF AI SYSTEMS IN LINE WITH THE EU ARTIFICIAL INTELLIGENCE ACT  
(OXFORD UNIVERSITY)

**capAI** [10] è una valutazione dell’impatto algoritmico pubblicata nel 2022 dalla *Oxford University*, sulla scia del testo di bozza dell’AI Act allora diffuso dalla Commissione Europea.

La valutazione dell’impatto è composta principalmente da tre documenti:

1. Protocollo di revisione interna (*Internal review protocol* - IRP): fornisce alle organizzazioni uno strumento di gestione del rischio e di garanzia della qualità;
2. Scheda di sintesi (*Summary Datasheet* - SDS) da presentare alla banca dati dell’UE;
3. Scorecard esterna e opzionale (*External Scorecard* - ESC).

Di particolare interesse ai fini di questo elaborato è il primo documento, l’**Internal Review Protocol**, che segue le cinque **fasi di sviluppo del ciclo di vita** di un sistema di IA e aiuta le organizzazioni ad essere consapevoli dei rischi del sistema e a mettere in atto adeguati processi di mitigazione.

- **Fase 1 “Design”**
  - Definizione della governance organizzativa;
  - Definizione dei requisiti funzionali al caso d’uso del sistema.
- **Fase 2 “Development”**
  - Raccolta di dati di buona qualità;

- Valutazione del modello.

### **Fase 3 “Evaluation”**

- Le informazioni richieste in questa fase riguardano:
  - la fase di test del sistema;
  - la fase di distribuzione del sistema.

### **Fase 4 “Operation”**

- Identificazione dei rischi associati alla modifica dei dati e al decadimento del modello;
- Descrizione del processo di reclamo dell'utente del sistema;
- Documentazione degli aggiornamenti del sistema.
- Rivalutazione periodica del sistema in relazione ai valori etici.

### **Fase 5 “Retirement”**

- Valutazione dei rischi di disattivazione del sistema;
- Descrizione documentata della strategia di gestione dei rischi per lo smantellamento del sistema.

La scorecard esterna (opzionale) richiede informazioni sullo scopo del sistema, sui valori e le norme organizzative alla base dello sviluppo del sistema, sui dati, sulla governance.

Nato in un **contesto accademico**, lo strumento si distingue per l'alto livello di dettaglio delle azioni prescritte per ogni fase e per l'identificazione dei ruoli coinvolti nella valutazione.

## AUDIT FRAMEWORK FOR ALGORITHMS (NETHERLANDS COURT OF AUDIT)

L'**Audit Framework for Algorithms** [11] è una valutazione dell'impatto algoritmico pubblicata nel 2021 dalla Corte dei conti dei Paesi Bassi.

L'Audit Framework for Algorithms è uno strumento pratico, implementato su fogli di lavoro Excel, ideato per aiutare le organizzazioni governative e del settore privato sia a identificare e mitigare eventuali rischi derivati dall'utilizzo dei sistemi algoritmici, sia a valutare se gli algoritmi utilizzati soddisfano criteri di qualità. Vista la sua praticità permette di adattare la valutazione al contesto di utilizzo dell'algoritmo.

Lo strumento si basa sulle "Ethics Guidelines for Trustworthy Artificial Intelligence", emanate dall'AI HLEG europeo.

La valutazione è basata principalmente su cinque aspetti:

1. **Governance and accountability.** Le otto domande trattano: la definizione di scopi, ruoli, responsabilità e competenze; la gestione del ciclo di vita dell'algoritmo; i fattori di rischio nell'uso dell'algoritmo; il coinvolgimento degli stakeholders esterni. Per ogni domanda, sono suggerite le misure di mitigazione da attuare.
2. **Model and data.** Le ventiquattro domande trattano: la qualità dei dati; lo sviluppo, l'uso e la manutenzione del modello algoritmico; la presenza di pregiudizi all'interno del dataset (dal punto di vista etico); la minimizzazione dei dati; eventuali test sul modello. Per ogni domanda, sono suggerite le mi-

sure di mitigazione da attuare.

3. **Privacy.** Le quindici domande analizzano se gli algoritmi utilizzati sono conformi al GDPR e, per ognuna, suggerisce le misure di mitigazione da mettere in atto.
4. **IT general controls (ITGC).** Le sedici domande riguardanti questo aspetto indagano sui controlli adottati dalle organizzazioni per garantire che i loro sistemi IT siano affidabili ed etici e, in particolare, si concentrano su: registrazione dei dati; diritti di accesso all'algoritmo; gestione delle password. Per ogni quesito sono poi suggerite le misure di mitigazione da attuare.
5. **Ethics.** Nonostante nel framework non sia presente una parte esplicitamente e specificamente dedicata a questo aspetto, l'eticità è concatenata e presente nelle domande poste per indagare tutti gli aspetti sopracitati. I principi etici cui l'Audit Framework for Algorithms si ispira sono i seguenti:
  - Rispetto dell'**autonomia umana**;
  - **Prevenzione dei danni** (principio diviso in: prevedibilità dell'algoritmo; protezione dei dati);
  - **Equità** (principio diviso in: non discriminazione; necessità di effettuare una valutazione dell'impatto sociale e ambientale);
  - **Spiegabilità e trasparenza** (principio diviso in: spiegazione sul funzionamento dell'algoritmo; spiegazione sulle procedure che sono state adottate).

Considerando che le domande rilevanti per effettuare una valutazione dipendono molto dal **contesto d'utilizzo dell'algoritmo** da valutare, gli autori suggeriscono agli auditor di selezionare le domande più utili in base al proprio contesto. Il framework, infatti, consente di filtrare sia per argomento sia per tipologia di rischio (anche in base a uno dei principi etici). Nella valutazione, per ogni domanda viene richiesto sia di fornire una risposta, sia di indicare il livello di rischio associato all'algoritmo, seppur il framework **non fornisca metriche o altre spiegazioni per poter giungere a conclusioni sull'impatto e sul rischio** del proprio sistema.

Una volta terminata la valutazione viene **consigliato di redigere una relazione** includendo eventuali miglioramenti e misure di mitigazione.

MODEL RULES ON IMPACT ASSESSMENT OF  
AUTOMATED DECISION-MAKING SYSTEMS USED BY  
PUBLIC ADMINISTRATION (EUROPEAN LAW INSTITUTE)

Il “**Model rules on impact assessment of automated decision-making systems used by public administration**” [12] è un documento pubblicato nel 2022 dall'**European Law Institute (ELI)**. ELI è un'organizzazione indipendente che ha lo scopo di avviare, condurre e facilitare la ricerca, formulare raccomandazioni e fornire guide pratiche nel contesto giuridico europeo. La sua proposta consiste in una valutazione dell'impatto dei sistemi decisionali automatizzati utilizzati all'interno della **pubblica amministrazione**.

Le Model Rules sono state concepite in modo da non dipendere dal diritto dell'UE e da poter essere **applicate anche in ordinamenti giuridici non appartenenti all'UE**. Le norme proposte sono state inoltre sviluppate in modo da garantire la loro compatibilità anche con la legislazione UE in fase di elaborazione: il riferimento è in particolare all'AI Act, il cui testo era ancora in bozza nel 2022.

La valutazione è composta da:

1. Un **questionario di screening** utile a determinare il **rischio del sistema** (vengono proposte alcune domande a risposta multipla e, in base alla risposta, si deve attribuire il punteggio indicato).
2. Una valutazione dell'impatto che aiuta nella redazione di una **relazione sull'impatto del sistema decisionale automatizzato**. Nella valutazione vengono proposte trentasei domande su: descrizione dello scopo e del funzionamento del sistema; pre-



stazioni, efficacia e efficienza del sistema rispetto agli obiettivi per cui viene implementato; diritti fondamentali, benessere sociale e ambientale; autonomia umana; misure adottate per garantire la massimizzazione dei benefici e la minimizzazione dei rischi identificati; supervisione umana e controllo del sistema; accuratezza, precisione, robustezza tecnica e sicurezza; trasparenza e spiegabilità; responsabilità e verificabilità.

Il framework suggerisce misure di mitigazione dei rischi emersi e identifica i ruoli coinvolti nella valutazione.

Viene proposta anche una versione estesa della valutazione, composta da sessanta domande che aiutano nella redazione di una relazione più dettagliata. Secondo l'ELI la relazione, una volta completata, deve essere **verificata da auditor esterni e indipendenti**.

La relazione, di cui è consigliata la revisione ogni sei mesi, dovrebbe essere pubblicata e resa consultabile online.

## FUNDAMENTAL RIGHTS AND ALGORITHM IMPACT ASSESSMENT - FRAIA (NETHERLANDS COURT OF AUDIT)

Il **Fundamental rights and algorithm impact assessment (FRAIA)** [13] è una valutazione dell'impatto algoritmico e dei diritti fondamentali pubblicata nel **2022** dalla Corte dei conti dei Paesi Bassi.

FRAIA è uno strumento pratico che comprende cinquantasette domande aperte alle quali le organizzazioni governative devono rispondere ogni volta che desiderano sviluppare, far sviluppare a terzi, acquistare o utilizzare un algoritmo. È consigliata la compilazione dello strumento ad un gruppo multidisciplinare (per ogni domanda, infatti, viene indicato chi dovrebbe essere coinvolto nella sua riflessione e compilazione).

FRAIA si basa sulle "Ethics Guidelines for Trustworthy Artificial Intelligence" dell'AI HLEG europeo. È composto da quattro parti:

1. **Part 1: why? Intended effects, objectives, preconditions.** Questa parte comprende otto domande relative a: la definizione del problema; lo scopo dell'algoritmo; i valori e i principi presi in considerazione nello sviluppo e nell'uso dell'algoritmo (compresi eventuali conflitti tra valori); le basi giuridiche di riferimento; gli stakeholders coinvolti (compreso l'indicare chi detiene la responsabilità).
2. **Part 2a: What? Data and input.** Questa parte comprende otto domande relative a: il tipo di algoritmo che verrà utilizzato; le fonti e la qualità dei dati; i pregiudizi inclusi nel dataset; la sicurezza dei dati.  
**Part 2b: What? Algorithm - throughput.** Questa

parte comprende quindici domande relative a: il tipo di algoritmo che verrà utilizzato; la proprietà dell'algoritmo; l'accuratezza dell'algoritmo; la spiegabilità e la trasparenza dell'algoritmo.

3. **Part 3: How? Implementation, use and supervision - output.** Questa parte comprende diciannove domande relative a: le decisioni basate sull'output dell'algoritmo; le decisioni finali prese dagli esseri umani sulla base dell'output algoritmico; gli effetti dell'algoritmo sugli stakeholder; le procedure messe in atto per soddisfare i requisiti di governance; il contesto (compresa la gestione dell'uso dell'algoritmo in un contesto diverso o cambiamenti nel contesto in cui opera l'algoritmo); la comunicazione (compresa l'apertura di dati e algoritmi); gli strumenti di valutazione, audit e tutela dell'algoritmo.
4. **Part 4: Fundamental rights.** Questa parte include una "tabella di marcia per i diritti fondamentali" composta da sette domande che hanno lo scopo di aiutare a identificare se l'algoritmo utilizzato inciderà sui diritti fondamentali e di facilitare la discussione al fine di prevenire o attenuare le eventuali ingerenze su questi diritti.

Infine, ogni parte del FRAIA comprende una spiegazione, le domande, il ruolo richiesto per rispondere, uno spazio per rispondere, gli aspetti chiave da considerare e le scelte fatte. FRAIA, quindi, può anche essere utilizzato come materiale di riferimento per l'accountability sul processo decisionale riguardante lo sviluppo e l'attuazione dell'algoritmo.

Lo strumento si propone come una valutazione completa, in quanto include anche altre linee guida, norme, principi nonché valutazioni dell'impatto (come la valutazione dell'impatto sulla protezione dei dati o l'Audit Framework for Algorithms sopracitato).

## ETHICAL IMPACT ASSESSMENT (UNESCO)

Nel 2023, l'**UNESCO** ha pubblicato l'**Ethical Impact Assessment** (EIA) [14] che ha la funzione di valutare l'eticità di un sistema di IA sulla base delle proprie Raccomandazioni sull'etica dell'intelligenza artificiale (UNESCO 2021).

Il framework di valutazione **incorpora considerazioni di tipo etico in tutto il processo di valutazione**. Esso si basa sul lavoro esistente prodotto da altre organizzazioni note: dove applicabile, alcune parti della valutazione prendono in prestito contenuti da strumenti o metodi esistenti, facendo loro espresso riferimento e accreditandoli nel processo. L'obiettivo è sfruttare la conoscenza preesistente, colmare le lacune esistenti e creare una metodologia coesiva che enfatizzi l'etica in tutti i passaggi.

L'EIA si compone di due parti principali:

**Parte 1 – Domande di inquadramento:** esse valutano gli aspetti fondamentali del progetto di intelligenza artificiale e, in base alle risposte date e alla luce delle linee guida stabilite nella Raccomandazione dell'UNESCO, si comprende se si può proseguire con la valutazione proposta dal framework.

La sezione intende stabilire solide basi etiche per il processo di acquisto o sviluppo dell'IA e si compone di 4 sezioni, proposte al team di progetto:

1. **Descrizione del progetto:** si specificano i dettagli del sistema che prevede di implementare e l'obiettivo che spera di raggiungere. Ciò aiuta a chiarire i dettagli del sistema e l'ambito dell'EIA. Aiuta anche a garantire che il piano sia ben adatto a un pro-

blema o a una sfida del mondo reale.

2. **Screening della proporzionalità:** la Raccomandazione dell'UNESCO contiene alcune disposizioni fondamentali che dovrebbero essere stabilite prima di passare alle fasi successive della valutazione d'impatto. Nello specifico, il sistema non dovrebbe essere utilizzato per la valutazione sociale o la sorveglianza di massa e la scelta di utilizzare l'IA deve essere giustificata in base al fatto che il metodo è proporzionale al raggiungimento dello scopo dichiarato.
3. **Governance del progetto:** per garantire un'efficace EIA, i ruoli e le responsabilità devono essere definiti fin dall'inizio. Deve essere garantita la diversità del team di progetto, poiché avrà un impatto sulla qualità del prodotto finale.
4. **Governance multi-stakeholder:** devono essere stabiliti anche piani per garantire che prospettive diverse siano incluse nell'EIA, incorporando i punti di vista e le esperienze di coloro che saranno impattati dal sistema.

**Parte 2 – Implementazione dei principi UNESCO:** questa sezione valuta se la progettazione, lo sviluppo e l'implementazione del sistema di IA risulteranno in processi e risultati coerenti con i principi etici dell'UNESCO per l'IA. Per ciascun principio, le domande mirano a valutare:

- a. Se sono state messe in atto **sufficienti salvaguardie procedurali** per garantire che questo sistema sia conforme alla Raccomandazione.

b. Gli **esiti positivi e gli impatti negativi** che possono o potrebbero derivare dall'acquisizione e dall'implementazione del sistema, specifici per il suo **contesto di utilizzo**.

A seguire, i principi dell'UNESCO a cui il sistema dovrebbe conformarsi:

- **Safety e Sicurezza;**
- **Correttezza, non-discriminazione, diversità;**
- **Sostenibilità;**
- **Privacy e Protezione dei dati personali;**
- **Trasparenza, Spiegabilità, Accountability e Responsabilità;**
- **Consapevolezza e Alfabetizzazione (*Literacy*).**

Infine, l'EIA si compone di un Annex che ha la funzione di individuare i potenziali rischi e impatti negativi e positivi legati al sistema di IA. Per ogni singolo impatto, ci sono due categorie da considerare: la *severity* (o il livello di significatività per i risultati positivi) e la *likelihood* (o probabilità).

Alla voce *severity* appartengono tre sottocategorie, ossia scala, portata e rimediabilità, che vanno a dettagliare e rendere più accurato il calcolo della *severity*.

## AI IMPACT ASSESSMENT: A TOOL TO SET UP RESPONSIBLE AI PROJECTS

Nel 2023, il **Ministero delle infrastrutture e della gestione delle acque olandese** [15] (*Ministry of Infrastructure and Water Management*) ha rilasciato il suo **AI Impact Assessment** per agevolare l'analisi dei sistemi basati sull'intelligenza artificiale in uso da parte delle **pubbliche amministrazioni olandesi**. Il tool si concentra sull'identificazione e l'analisi di eventuali ostacoli o problematiche che possono emergere a livello di dati, sistemi o algoritmi, assicurandosi che siano rispettate le normative e le leggi vigenti.

L'Impact Assessment agisce come un meccanismo per **dialogare e documentare il processo decisionale**, contribuendo a migliorare aspetti cruciali come la responsabilità, la qualità e la capacità di riproduzione dei risultati. L'obiettivo finale è la creazione di un **documento chiaro e facilmente comprensibile**, che illustri in maniera dettagliata le ragioni e i criteri che hanno guidato la scelta di impiegare l'IA all'interno di un determinato progetto.

È erogato sottoforma di **questionario** e si compone di **100 domande**, alcune obbligatorie, altre facoltative, altre ancora a cui si è tenuti a rispondere solo se si sta valutando un sistema ad alto rischio (scelta che ricalca la strada intrapresa dall'IA ACT europeo). La valutazione di impatto viene definita come "obbligatoria ma flessibile". Ciò significa che è lasciato principalmente al buon senso del produttore il valutare o no l'impatto del proprio sistema di Intelligenza Artificiale.

Le domande si dividono in 6 sezioni, qui di seguito



brevemente descritte.

La **sezione introduttiva** pone le basi per la discussione, focalizzandosi sugli **aspetti generali** del sistema di intelligenza artificiale (IA) in sviluppo. Si concentra principalmente su due aree: lo **scopo del sistema** e il suo **ruolo all'interno dell'organizzazione**. Questo include una riflessione sulle tecnologie impiegate, sui responsabili del progetto e su come il sistema si inserisce nella strategia complessiva dell'organizzazione.

Le domande della seconda sezione, invece, evidenziano l'importanza di **proteggere i diritti fondamentali** nell'uso dell'IA e di affrontare i rischi di violazione di tali diritti. Si sottolinea la necessità di prevenire i *bias* nei dati, nel modello e nei risultati, riconoscendo come questi possano rafforzare le disuguaglianze esistenti. Viene inoltre messa in evidenza l'importanza della partecipazione inclusiva e della diversità degli stakeholder in tutto il ciclo di vita del sistema di IA.

La terza sezione riguarda la **robustezza tecnologica** che viene esaminata attraverso vari aspetti, tra cui l'accuratezza, l'affidabilità, l'implementazione tecnica, la riproducibilità e la spiegabilità. Questi elementi contribuiscono a garantire che il sistema sia efficace, affidabile e comprensibile sia per gli sviluppatori che per gli utenti finali.

Un altro aspetto importante, oggetto di analisi nella quarta sezione, è la **gestione dei dati** utilizzati dai sistemi di IA, in particolare, l'accesso, la proprietà, l'utilizzabilità, l'integrità, la sicurezza e la qualità dei dati. Si rileva l'importanza della privacy e della protezione dei dati personali, conformemente a regolamenti come il

GDPR.

La quinta sezione si concentra sulla **gestione dei rischi** associati all'uso dei sistemi di IA, inclusi i rischi **imprevisti** che possono compromettere l'affidabilità del sistema e causare danni. Viene sottolineato il principio di prevenzione per minimizzare i danni.

Infine, la sezione sull'**accountability** affronta la necessità di stabilire meccanismi adeguati a garantire la responsabilità nell'uso dei sistemi di IA. Questo include comunicazioni chiare agli utenti, processi di verifica dei dati e dei modelli, e l'archiviazione delle informazioni relative al sistema. Viene anche trattato l'adattamento climatico, sottolineando l'importanza dell'uso ecologico dell'IA e il suo potenziale nel contribuire a soluzioni di problematiche ambientali.

Le sezioni di questa valutazione di impatto delineano un **quadro per lo sviluppo e l'implementazione responsabile di sistemi di intelligenza artificiale** nell'ambito della pubblica amministrazione, mettendo in primo piano considerazioni etiche, di governance, di robustezza tecnologica e di gestione dei rischi.

## AI GOVERNANCE FRAMEWORK FOR GENERATIVE AI (SINGAPORE)

Nel gennaio del 2024, la **Infocomm Media Development Authority** (IMDA) del Governo di Singapore insieme ad AI Verify Foundation ha pubblicato una **proposta di Modello di Governance per l'Intelligenza Artificiale Generativa** [16]. Il documento aggiorna il Modello di Governance dell'IA esistente già dal 2019.

Il Modello di Governance dell'IA per l'Intelligenza Artificiale Generativa mira a instaurare un metodo sistematico ed equilibrato per **gestire le problematiche** associate all'IA generativa, **promuovendo** al contempo **l'innovazione**. Questo approccio richiede una partecipazione attiva di tutti i principali attori coinvolti, riconoscendo ad ognuno un ruolo fondamentale: dai responsabili delle decisioni politiche, alle aziende, passando per la comunità scientifica fino ad arrivare al grande pubblico.

Con la finalità di creare un ambiente di fiducia nell'uso dell'IA generativa, il Framework suggerisce l'esame di **nove aspetti fondamentali**, considerandoli in un'ottica complessiva:

- **Accountability**. Incoraggiare chi sviluppa IA a essere responsabile verso gli utenti finali è un passo fondamentale per costruire la fiducia, seppur nella difficoltà di capire a chi attribuire la responsabilità in un contesto stratificato e complesso come quello dell'IA di tipo generativo. Anche se quest'ultima ha caratteristiche uniche, è possibile trovare paralleli utili con gli attuali sistemi di cloud e sviluppo software, e intraprendere i primi passi pratici nell'i-

identificazioni di ruoli e responsabilità.

- **Dati.** I dati sono essenziali nello sviluppo di modelli IA e influenzano notevolmente la qualità dei risultati. È importante assicurarsi che i dati con cui si alimenta il modello siano di alta qualità, ad esempio utilizzando fonti affidabili. Nei casi in cui l'uso dei dati per l'addestramento del modello possa essere controverso, come nel caso di dati personali o materiali coperti da proprietà intellettuale, è importante garantire chiarezza commerciale, trattamento equo e agire in modo pragmatico.
- **Sviluppo e Implementazione affidabili.** Lo sviluppo del modello e la sua implementazione sono fondamentali per l'innovazione guidata dall'IA. Nonostante la visibilità limitata degli utenti finali, è essenziale la trasparenza riguardo le misure di sicurezza adottate. Questo coinvolge l'adozione delle migliori pratiche nell'industria per lo sviluppo, la valutazione e la trasparenza e divulgazione simile all'etichetta degli alimenti.
- **Segnalazione di Incidenti.** Anche con i processi di sviluppo più robusti, nessun software è completamente infallibile, e lo stesso vale per l'IA. La segnalazione di incidenti permette una notifica e un rimedio tempestivi. È quindi fondamentale stabilire strutture e processi per monitorare e segnalare incidenti, supportando il miglioramento continuo dei sistemi IA.
- **Test e Garanzia.** Nel contesto di un ecosistema di fiducia, i test e le garanzie di terze parti giocano un ruolo complementare. Sebbene per l'IA questo sia

ancora un campo emergente, è utile adottare test e garanzie di terze parti per dimostrare la fiducia agli utenti finali. Sviluppare standard comuni per i test dell'IA è anche importante per garantire qualità e coerenza.

- **Sicurezza.** L'IA generativa introduce potenziali nuove minacce attraverso i modelli stessi, superando rischi e minacce note per gli altri sistemi software. I framework esistenti per la sicurezza delle informazioni devono essere adattati al nuovo contesto e devono essere sviluppati nuovi strumenti di test per affrontare questi rischi.
- **Provenienza dei Contenuti.** I contenuti generati dall'IA, a causa della facilità con cui possono essere creati, possono esacerbare la disinformazione. La trasparenza su dove e come i contenuti sono generati permette agli utenti finali di consumare i contenuti online in modo informato. I governi stanno esplorando soluzioni tecniche come la filigrana digitale e la provenienza crittografica.
- **Ricerca e Sviluppo per la sicurezza e l'allineamento.** L'attuale stato della scienza per la sicurezza dei modelli non copre tutti i rischi. È necessario un investimento accelerato in ricerca e sviluppo per migliorare l'allineamento dei modelli con le intenzioni e i valori umani.
- **IA per il bene pubblico.** Fare IA responsabilmente non significa soltanto mitigare i rischi, ma anche elevare e dare potere alle persone e alle aziende per prosperare in un futuro abilitato dall'IA. Democratizzare l'accesso all'IA, migliorare l'adozione

dell'IA nel settore pubblico, formare i lavoratori e sviluppare sistemi di IA in modo sostenibile sosterrà gli sforzi per indirizzare l'IA verso il bene pubblico.

Uno dei primi tentativi di disciplinare il fenomeno dirompente dell'IA generativa, questo documento è da considerarsi una bozza ed è aperto a opinioni, commenti e feedback da parte della comunità internazionale fino a metà 2024.

## ARTIFICIAL INTELLIGENCE RISK MANAGEMENT FRAMEWORK (NIST)

**L'Artificial Intelligence Risk Management Framework (AI RMF)** [17] è stato realizzato dal National Institute of Standards and Technology (NIST) nel gennaio 2023. Seppur non strettamente rientrante nella categoria delle valutazioni di impatto, si ritiene importante citare questo framework sia per la autorevolezza della fonte sia per l'intento di fornire alle organizzazioni un **sistema integrato per la gestione dei rischi collegati all'IA**.

Il Framework è stato concepito per fornire a organizzazioni e individui uno strumento per **aumentare l'affidabilità dei sistemi** di IA e promuovere la progettazione, lo sviluppo, la distribuzione e l'uso responsabile dei sistemi di IA nel tempo. Esso è concepito per essere volontario, rispettoso dei diritti, non specifico per settore e agnostico all'uso, in modo tale che possa facilmente essere approcciato dalle organizzazioni di tutte le dimensioni e in tutti i settori e nella società.

Il Framework si divide in **due parti**: la **prima parte** illustra come le organizzazioni possono inquadrare i rischi legati all'IA, e individuare i destinatari del sistema; la **seconda parte, che** costituisce il "core" del Framework, descrive **quattro funzioni specifiche** per aiutare le organizzazioni ad affrontare i rischi dei sistemi di IA nella pratica. Ciascuna di queste funzioni di alto livello è suddivisa in categorie e sottocategorie, a loro volta suddivise in azioni specifiche e risultati. Le azioni non costituiscono una lista di controllo, né sono necessariamente un insieme ordinato di passaggi.

Le funzioni sono:

- **GOVERN**: è una funzione trasversale che si diffonde in tutta la gestione del rischio di IA e abilita le altre funzioni del processo. Gli aspetti di GOVERN, in particolare quelli relativi alla conformità o alla valutazione, dovrebbero essere integrati in ciascuna delle altre funzioni. L'attenzione alla governance è un **requisito continuo e intrinseco** per una gestione efficace dei rischi dell'IA per tutta la durata di vita di un sistema di IA e per la gerarchia dell'organizzazione.
- **MAP**: crea un **contesto** per capire i rischi associati a un sistema di IA. Questo sistema coinvolge molte attività connesse tra loro e vari attori. Coloro che gestiscono una parte del processo spesso non hanno un controllo completo sulle altre parti. Queste connessioni possono rendere difficile prevedere gli effetti dei sistemi di IA. Ad esempio, le decisioni prese all'inizio sullo scopo di un sistema di IA possono modificare il funzionamento del sistema stesso e i suoi impatti su chi lo usa. Quindi, anche se si hanno buone intenzioni in una parte del processo, queste possono essere compromesse da decisioni prese in altre fasi.
- **MEASURE**: impiega strumenti, tecniche e metodi quantitativi, qualitativi o misti, e metodologie per analizzare, valutare, confrontare e monitorare il rischio collegato all'IA e i relativi impatti. Utilizza le conoscenze relative ai rischi di IA identificati nella funzione MAP e informa la successiva funzione MANAGE. I sistemi di IA devono essere testati



prima dell'implementazione e regolarmente durante il funzionamento. Le misurazioni del rischio di IA includono la documentazione degli aspetti della funzionalità e dell'affidabilità dei sistemi.

**MANAGE:** prevede di destinare risorse per gestire i rischi identificati e valutati, come stabilito dalla funzione GOVERN. Questo include piani per affrontare, superare e comunicare eventuali incidenti. Le informazioni raccolte dagli esperti e dai principali attori coinvolti sono usate in questa fase per ridurre la possibilità di fallimenti del sistema e suoi impatti negativi. La documentazione sistematica aiuta nella gestione dei rischi AI, migliorando trasparenza e responsabilità. Sono previsti processi per identificare nuovi rischi e strategie per il miglioramento continuo. Dopo aver implementato la funzione MANAGE, si avranno piani per dare priorità ai rischi e per il monitoraggio e miglioramento continuo. Questo permetterà agli utenti del Framework di gestire meglio i rischi dei sistemi AI e di assegnare risorse in modo efficace. Gli utenti del Framework devono continuare ad applicare questa funzione ai sistemi di IA in uso, adattandosi a cambiamenti nei metodi, contesti, rischi e nelle esigenze degli attori coinvolti.

Il documento è completato da una serie di **Appendici:** A) descrive gli attori che sono coinvolti nei processi di IA; B) evidenzia le differenze tra i sistemi di IA ad alto rischio rispetto ai software ad alto rischio; C) si occupa del risk management e dell'interazione del sistema di IA con l'essere umano; D) descrive in breve le caratteristiche della valutazione.



## PARTE III

### RIFLESSIONI ED ESPERIENZE



## RIFLESSIONI: HUMAN OVERSIGHT

Il nuovo Regolamento europeo in tema di intelligenza artificiale (IA) prescrive che il datore di lavoro/utilizzatore (*deployer*) di un sistema di IA ‘ad alto rischio’ sia tenuto a designare un lavoratore - subordinato od autonomo - al quale affidare l’attività di sorveglianza (*human oversight*) del funzionamento del sistema stesso (art. 26, par. 2).

Lo svolgimento di tale attività, a sua volta disciplinata in dettaglio dall’art. 14 del citato Regolamento, deve essere affidato ad un soggetto (*natural person*) che risulti essere:

- in possesso di una comprovata competenza professionale;
- opportunamente formato, anche con riferimento al rischio di fare eccessivo affidamento sui dati elaborati dal sistema (*automation bias*), specie quando tali elaborazioni sono impiegate a supporto di decisioni che devono essere assunte dagli organi di direzione dell’impresa;
- investito dell’autorità (*authority*) che consenta l’effettivo e compiuto svolgimento dell’attività di sorveglianza nell’ambito dell’organizzazione;
- dotato del supporto - oltre che di natura organizzativa, anche economico - necessario all’esercizio della propria funzione.

È altresì stabilito che l’attività di sorveglianza umana è svolta al fine di i) prevenire e ii) ridurre al minimo i rischi per la salute, la sicurezza o l’esercizio dei diritti

fondamentali che potrebbero presentarsi nell'ipotesi in cui un datore di lavoro impieghi un sistema di IA ad alto rischio anche in modo improprio, ma ragionevolmente prevedibile.

Non solo: fermo restando che le misure di sorveglianza debbono essere adeguate rispetto al contesto organizzativo nel quale il sistema di IA è utilizzato - nonché alla natura ed entità dei rischi individuati così come al grado di autonomia con il quale si è stabilito operi il sistema di AI - è fondamentale evidenziare che l'ingegnerizzazione di un sistema deve essere sviluppata dal *provider* in modo da consentire e facilitare la sorveglianza da parte del soggetto designato dal datore di lavoro/utilizzatore.

In sostanza, il sistema di AI deve essere progettato e sviluppato in modo tale da consentire un i) efficace e ii) continuo controllo circa il proprio funzionamento, offrendo adeguati strumenti di interfaccia alla persona fisica incaricata di svolgere l'attività di sorveglianza.

Quanto precede implica che il soggetto deputato allo svolgimento dei compiti di sorveglianza deve poter:

- comprendere correttamente le capacità e i limiti del sistema in uso;
- monitorare il funzionamento del sistema e tempestivamente rilevare eventuali anomalie così come disfunzioni e prestazioni inattese;
- interpretare correttamente l'elaborato dal sistema, facendo ricorso a strumenti e metodi d'interpretazione indicati sia dal *provider* così come dal datore di lavoro/utilizzatore;

- decidere, al ricorrere di una situazione particolare, di non usare il sistema ovvero ignorarne o annullarne l'elaborato;
- intervenire sul funzionamento del sistema sino a poterne disporre l'arresto qualora si presentino anomalie che potrebbero recare pregiudizio per la salute, la sicurezza o i diritti fondamentali dei lavoratori in forza.

Alla luce di quanto sopra, è evidente come la scelta da parte di un datore di lavoro/utilizzatore di avvalersi di un sistema di IA ad alto rischio debba essere fondata anche su un esame (documentato):

- delle funzionalità che il *provider* fornisce per lo svolgimento dell'attività di sorveglianza del sistema nonché dei criteri e degli strumenti d'interpretazione dell'*output*;
- dell'attività di addestramento del sistema e della natura dei dati impiegati a tal fine;
- della trasparenza del sistema;
- delle informazioni rese in osservanza dell'art. 13, par. 2 del Regolamento, per effetto del quale il *provider* è tenuto a fornire appropriate istruzioni per l'uso del sistema che siano i) concise, ii) complete, iii) corrette, iv) pertinenti, v) accessibili e vi) comprensibili.

Data la funzione cruciale che la legislazione comunitaria attribuisce all'attività di sorveglianza umana di un sistema di IA ad alto rischio, è altresì opportuno svolgere alcune considerazioni di ordine organizzativo:

- il datore di lavoro deve responsabilmente individuare una persona fisica in possesso dei requisiti professionali e formativi, ma anche di caratteristiche soggettive che consentano l'effettivo svolgimento del ruolo all'interno dell'organizzazione aziendale;
- nell'atto di formale designazione predisposto dal datore di lavoro/committente, e sottoscritto dalle parti, siano indicati in dettaglio:
- i compiti e le dotazioni strumentali, anche finanziarie, attribuite alla funzione,
- termini, modalità e procedure per lo svolgimento dell'attività;
- i poteri d'intervento attribuiti al soggetto designato;
- le procedure implementate per l'adempimento degli obblighi di comunicazione che la legislazione vigente prescrive o la cui effettuazione sia prevista dal sistema di gestione adottato dal datore di lavoro/committente;
- nel caso in cui l'attività di sorveglianza del sistema sia giudicata particolarmente complessa e richieda il concorso di più professionalità, è possibile che il datore di lavoro/committente individui più persone fisiche, introducendo un sistema di regole che consentano a tale organo collegiale di esercitare la propria funzione in modo efficace e tempestivo.

Sul piano organizzativo, si consideri peraltro il rilievo che assume l'obbligo posto in capo al datore di lavoro/utilizzatore di comunicare al *provider* eventuali malfunzionamenti del sistema (art. 26, par. 5); infat-



ti, ancorché tale obbligo sia espressamente attribuito al datore di lavoro/utilizzatore, il suo adempimento non può che esser promosso su impulso dell'addetto alla sorveglianza, proprio in ragione del ruolo e delle responsabilità attribuitigli dalla normativa vigente nonché dall'atto di designazione mediante il quale ha accettato l'incarico di svolgere l'attività di sorveglianza. In considerazione del fatto che un rilievo negativo o una grave anomalia del funzionamento dell'applicativo di AI potrebbe comportare il suo arresto (*shut down*), è imprescindibile l'individuazione di una figura professionale che sia altresì in grado di valutare in modo appropriato le conseguenze derivanti da tale decisione.

La legislazione comunitaria delinea dunque una nuova figura professionale a cui attribuire la responsabilità diretta del corretto funzionamento del sistema di IA. Pertanto, la responsabilità della funzione di *human oversight* è necessaria sia affidata ad una persona dotata di autonomia, indipendenza di valutazione e autorevolezza, requisiti che inducono a propendere per l'individuazione di un lavoratore autonomo che, sottratto ai condizionamenti che potrebbero caratterizzare un rapporto di lavoro subordinato, sia in ogni caso attento a preservare le esigenze economiche e di produzione nel rispetto dei valori etici disposti dagli ordinamenti lavoristici.

*A cura di ArlatiGhislandi*



## **ESPERIENZE: L'APPROCCIO DI HAIER EUROPE PER UN USO RESPONSABILE DELL'INTELLIGENZA ARTIFICIALE**

Haier Europe ha una lunga storia nell'IoT, iniziata nel 2014 con la prima gamma di elettrodomestici connessi (Simply-Fi e Wizard). Il gruppo non ha mai smesso di lavorare sull'innovazione e sull'espansione dell'ecosistema connesso. In Europa, la soluzione multibrand hOn, lanciata sul mercato nel 2020, vanta attualmente il controllo di 27 linee di prodotto (Elettrodomestici, Climatizzazione, Fotovoltaico) con oltre 5.000 prodotti ed è disponibile in 25 lingue. hOn, sin dal lancio, gode di un alto livello di apprezzamento da parte degli utenti, frutto di un continuo affinamento e di un livello di innovazione elevato che hanno reso la Soluzione Smart di Haier Leader nelle Smart Appliances, con un rating medio sugli App Store di 4.4. Tutte le funzioni che offriamo, seguendo l'approccio 'Zero Distance to Consumer', sono progettate per rispondere a esigenze reali e comuni. Il nostro obiettivo è offrire un'esperienza di vita intelligente e completa che ti segua e ti aiuti nella quotidianità

Nel panorama sempre più complesso e affascinante dell'Intelligenza Artificiale, Haier Europe si erge come pioniere, spingendo i confini dell'innovazione e trasformando radicalmente settori tradizionali come quello degli elettrodomestici. Abbiamo iniziato a sviluppare prodotti e soluzioni basate su A.I. nel 2016 e da allora abbiamo progressivamen-

te allargato lo spettro di utilizzo attraverso hOn su tutte le famiglie di prodotto, creando un'esperienza complessiva utile, performante e facile da utilizzare. Grazie all'Intelligenza Artificiale, applicata in modo innovativo alle varie categorie di prodotto, è possibile garantire risultati personalizzati per il singolo consumatore, eliminando la possibilità di errore e rendendo più efficiente l'utilizzo dei nostri prodotti attraverso hOn. Elementi come sostenibilità, durabilità e facilità d'uso sono la base di ogni nuova soluzione che introduciamo sul mercato. Un altro importante ambito su cui operiamo è garantire affidabilità e performance durature nel tempo, attraverso funzionalità di manutenzione preventiva e supporto al cliente tramite servizi di assistenza proattivi.

Uno dei vertici di questa innovazione è la gamma di forni dotati di una telecamera integrata che grazie all'A.I. hanno la funzione di cottura automatica, ovvero riconoscono il cibo e gestiscono la cottura e la ricetta con parametri ottimali e specifici per il cibo immesso. Questo non solo migliora l'esperienza dell'utente, ma ottimizza anche il processo di cottura, garantendo risultati eccellenti ogni volta.

L'integrazione dell'A.I. comporta inevitabilmente la raccolta e l'elaborazione di grandi quantità di dati personali. Haier Europe adotta politiche di sicurezza dei dati stringenti e trasparenti per proteggere la privacy degli utenti, seguendo un approccio "by design" che incorpora la protezione dei dati fin dalla fase iniziale di progettazione. Ciò significa che ogni aspetto dello

sviluppo tecnologico e del ciclo di vita del prodotto considera la privacy e la sicurezza dei dati come priorità fondamentali.

Inoltre, consapevole del crescente interesse dei consumatori, l'azienda ha intrapreso un percorso di implementazione di sistemi di sicurezza e certificazioni che superano i requisiti delle attuali normative Europee, anticipando gli standard che diventeranno obbligatori nel prossimo futuro.

È evidente che l'Intelligenza Artificiale sollevi numerose questioni etiche che richiedono un'attenta considerazione da parte dell'azienda e degli operatori coinvolti nell'intero percorso. Haier è particolarmente attenta ed attiva su questi temi che vengono gestiti fin dallo stato embrionale del progetto, attenzionando con particolare cura aspetti di Trasparenza, Responsabilità, Privacy, Sicurezza e Discriminazione.

L'Intelligenza Artificiale offre immense opportunità per innovare e migliorare il settore degli elettrodomestici. Tuttavia, è fondamentale che queste tecnologie siano implementate con un approccio responsabile alla gestione dei dati e che il relativo output, che in primis, sia basato sui principi di etica che da sempre contraddistinguono Haier.

Haier Europe si impegna a rispettare i principi di data protection, garantendo che l'innovazione tecnologica vada di pari passo con la sicurezza, la privacy ed i principi di etica. In questo modo, l'azienda non solo migliora la vita quotidiana dei consumatori, attraverso prodotti e soluzioni altamente performanti, ma costruisce anche un futuro più sicuro e trasparente.



PARTE IV

CONCLUSIONI





## **OLTRE UNA VISIONE TECNOCENTRICA**

Negli ultimi decenni, l'**innovazione**, soprattutto di natura tecnologica, è diventata la “stella polare” che orienta, in modo trasversale, governi, istituzioni internazionali, imprese private, comunità scientifica e società civile – intesa nel suo complesso (Lepore 2014; Godin 2015). La retorica che accompagna tali processi ha trasformato l'espressione “innovazione negativa” in un ossimoro. Ai giorni nostri, le persone si dicono quasi sempre favorevoli all'innovazione: la nozione, infatti, ha una connotazione positiva, visto che risulta capace di trasmettere qualcosa di moderno e affascinante, che crea un punto di discontinuità con il passato – per definizione, da superare – e che, pertanto, deve essere sostenuto senza remore (Verzelloni 2020a).

Eppure, anche l'innovazione ha un “**lato oscuro**”: essa può generare degli effetti perversi, inattesi e paradossali, di diversa natura e portata, che possono manifestarsi a vari livelli – individuale, intra-organizzativo, organizzativo, inter-organizzativo e sistemico. Se non adeguatamente compresi e governati, questi processi di innovazione possono dare vita a una serie di distorsioni, diseconomie e criticità, come, per esempio: rallentare l'attività lavorativa, disincentivare creatività e impegno delle persone, ledere il clima e diminuire il benessere organizzativo, favorire l'aumento di stress lavoro-correlato e l'emersione di fenomeni di burnout e compromettere, anche irrimediabilmente, la reputazione esterna di un'impresa e il suo business.

Questo discorso risulta particolarmente evidente – nonché di grande attualità e salienza – con riferimento alle innovazioni tecnologiche e, in particolare, allo sviluppo e all'implementazione di soluzioni basate su sistemi di **intelligenza artificiale** – d'ora in poi, anche AI. Al di là dei luoghi comuni, tali strumenti non possono essere concepiti al pari di un software da installare nei computer dei dipendenti o da dispiegare nelle diverse filiali di un'impresa. Per comprendere adeguatamente i loro effetti di natura sistemica, è necessario fare ricorso a specifiche lenti teoriche e metodologiche e investire, al contempo, nella promozione di percorsi di ricerca, formazione, assistenza e consulenza.

Come emerso nel corso dell'evoluzione della letteratura internazionale su questi temi, soprattutto di stampo sociologico, ma non solo, occorre ripensare radicalmente il modo stesso di intendere il ruolo delle tecnologie. Tali strumenti non sono e non possono essere una **tecnologia-in-sé**, ma devono essere concepiti come una specifica **tecnologia-in-uso** (Suchman 1987; Latour 1988; Gherardi 2009). Lungi dall'essere elementi neutri, le tecnologie possono essere paragonate a degli “attori non-umani”, che si inseriscono, influenzandolo, in un campo strutturato di pratiche lavorative. A loro volta, essendo coinvolte in un processo continuo di “mediazione pratica” anche le tecnologie possono trasformarsi ed essere utilizzate dalle persone per raggiungere obiettivi e per svolgere attività anche molto diverse da quelle previste in fase di progettazione e sviluppo (Bijker et al. 1987; Bijker 1999; Piana e Verzelloni 2022).

Questa concezione rappresenta una critica manifesta alla “visione tecnocentrica” ancora imperante in molti ambiti del sapere, anche non strettamente informatici e ingegneristici. Come rilevato da Norman (1998), il tecnocentrismo si fonda sull’assunto che esista una relazione di tipo diretto tra incremento e progressivo miglioramento delle tecnologie – di diverso tipo e portata – e innalzamento dei livelli di efficacia, efficienza, sicurezza e affidabilità. Come illustra la tabella sottostante (Tab. 1), la “**visione tecnocentrica**” propone, di fatto, una contrapposizione tra persone e macchine. Secondo tale approccio, infatti, le persone dovrebbero essere destinate a svolgere operazioni residuali, visto che, a differenza delle macchine, sono per propria stessa natura: vaghe, disorganizzate, soggette a distrazioni, emotive e illogiche (Norman 1993, 1998).

PERSONE	MACCHINE
Vaghe	Precise
Disorganizzate	Ordinate
Soggette a distrazioni	Non soggette a distrazioni
Emotive	Non emotive
Illogiche	Logiche

*Tab. 1. La visione centrata sulla macchina.*

*Fonte: Norman 1998, tratto da Catino 2006: 14.*

Ma è davvero così? Siamo davvero sicuri che adottando nuove tecnologie, magari sempre più sofisticate e complesse, otterremmo sicuramente, in tempi brevi, un aumento dei livelli di efficacia, efficienza, sicurezza

e affidabilità dei contesti in cui questi stessi strumenti si inseriscono? Esiste davvero una relazione di causa-effetto tra ricorso all'innovazione – rispetto ai nostri fini, di natura tecnologica – e performance organizzative?

Questo dibattito ha origini lontane, che possono essere fatte risalire all'emergere dalla contrapposizione tra **democrazia e tecnocrazia** (Verzelli 2020b) e, in particolare, alla diffusione dei principi dell'organizzazione scientifica del lavoro. In questo senso, il pensiero di Taylor è stato rivoluzionario, dal momento che si fondava sull'assunto che, attraverso l'organizzazione sistematica dell'attività umana, fosse possibile eliminare qualsiasi inefficienza. È in questo scenario che un ingegnere californiano, William H. Smyth utilizza, per la prima volta, il termine tecnocrazia e che autori come Veblen e Bellamy ispirano, attraverso le loro opere, la nascita del movimento tecnocratico.

Dopo la Seconda guerra mondiale, con il rapido sviluppo delle conoscenze scientifiche, tecniche e tecnologiche, in quella che viene definita come una “epoca d'oro della tecnocrazia”, i tecnocrati entrano stabilmente nelle organizzazioni pubbliche e private, come testimonia la progressiva diffusione delle c.d. “tecnostrutture” (Galbraith 1967). Nelle grandi corporation americane – ma non solo – il potere e, di conseguenza, la capacità di incidere sulle dinamiche economiche e sociali, passa, idealmente, delle mani dell'imprenditore a quelle di coloro che per competenze ed esperienza rappresentano l'intelligenza-guida o il cervello di un'organizzazione (Verzelli 2020b).

Eppure, nonostante la retorica dei presunti successi della tecnocrazia, numerose evidenze empiriche dimostrano che, se si adotta una “visione tecnocentrica”, non si può cogliere appieno la complessità che contraddistingue ogni contesto sociale e organizzativo. Il quadro è molto più articolato e impone di interrogarsi criticamente, in modo sistematico, sugli effetti, anche nel medio e lungo periodo, di ogni innovazione.

Come messo in luce da una parte pionieristica della letteratura organizzativa, infatti, l'esistenza di una connessione diretta tra **propensione all'innovazione** e **performance organizzative** risulta difficilmente verificabile, per almeno 5 ragioni (Verzelloni 2020a):

- non tutte le innovazioni si traducono in effettivi miglioramenti, così come non tutti i miglioramenti derivano da un'idea innovativa;
- i risultati collegati all'introduzione di un'innovazione possono anche manifestarsi dopo un intervallo di tempo, più o meno ampio;
- alcune innovazioni falliscono prima di essere implementate (Van de Ven 1986), ma possono comunque incidere sull'attività degli attori di un dato contesto organizzativo. Tuttavia, il bias positivo che contraddistingue questo settore, fa sì che, se un'innovazione non produce i risultati previsti, viene quasi sempre scartata, in quanto catalogata come un errore (Kimberly ed Evanisko 1981);
- innovare può produrre, inizialmente, un calo delle performance, determinato dai costi che devono essere necessariamente sostenuti per supportare lo sviluppo e l'implementazione dell'innovazione.

Inoltre, gli attori necessitano sempre di un periodo di tempo, più o meno ampio, per adattare e riformulare le loro pratiche lavorative e sperimentarne, eventualmente, di nuove (Corradi et al. 2010). Tutto ciò può incidere negativamente sia sul funzionamento sia sulle performance di un contesto organizzativo;

- individuare e quantificare l'impatto di un'innovazione può risultare oltremodo difficile, dal momento che l'introduzione di un'idea innovativa può produrre degli effetti anche molto diversi nei vari settori e nelle varie articolazioni di un sistema (Van de Ven 1986). A ciò si deve aggiungere che, da un lato, gli esiti di un percorso di innovazione possono essere influenzati da una pluralità di variabili e, dall'altro, che l'eventuale applicazione simultanea di diverse soluzioni, rende estremamente complicato distinguere l'impatto dei singoli interventi.

Sulla stessa linea, Neil Postman (2000: 42-53) ha proposto sette domande, che è necessario porsi – idealmente – prima di adottare una nuova tecnologia:

1. *Qual è il problema che questa nuova tecnologia risolve?*
2. *Di chi è il problema?*
3. *Quali nuovi problemi creiamo risolvendo questo problema?*
4. *Quali persone e istituzioni saranno maggiormente interessate da una soluzione tecnologica?*
5. *Quali cambiamenti linguistici si verificano a seguito del cambiamento tecnologico?*

6. *Quali cambiamenti nel potere economico e politico potrebbero derivare dall'adozione di questa tecnologia?*
7. *Quali usi alternativi (e non intenzionali) potrebbero essere fatti di questa tecnologia?*

Ancora una volta, pertanto, le tecnologie non possono essere concepite come degli elementi esogeni, che si integrano subito, in modo automatico, nelle pratiche lavorative delle persone (Verzelloni 2019) e nelle logiche di funzionamento organizzativo, risolvendone i problemi e migliorandone le performance. Anche le tecnologie sono, infatti, degli attori che hanno un impatto, più o meno significativo, sul contesto sociale e organizzativo in cui si inseriscono e che sono, a loro volta, modellate dall'attività delle persone.

Inserendosi in questo dibattito, è fondamentale per gli stessi attori organizzativi, poter autovalutare la propria consapevolezza sui possibili effetti derivanti dall'applicazione dell'intelligenza artificiale.

Per comprendere l'impatto di una tecnologia – come le soluzioni basate sui sistemi di intelligenza artificiale – occorre allargare lo sguardo, ossia ricostruire e analizzare il contesto sociale e organizzativo in cui si inserisce e prefigurare i suoi possibili effetti, ai diversi livelli.

Senza questo allargamento di prospettiva, si possono fornire ricostruzioni soltanto parziali – e, pertanto, fuorvianti – del reale impatto di una soluzione tecnologica sia sulle pratiche lavorative delle persone che operano in un dato contesto organizzativo sia, più in

generale, sui suoi potenziali utilizzatori e sulla società, intesa nel suo insieme.

*Luca Verzelloni*  
*ricercatore di Sociologia dell'organizzazione*  
*presso il Dipartimento di Sociologia e ricerca sociale*  
*dell'Università degli studi di Milano-Bicocca*



## L'APPROCCIO OLISTICO ALLA GOVERNANCE DELL'IA

Nel chiudere questa disamina, allarghiamo leggermente l'inquadratura per sottolineare che c'è un approccio responsabile e sostenibile all'introduzione delle tecnologie di IA. È quello di **integrare la valutazione di impatto algoritmico**, qualsiasi modello si segua per effettuarlo, **in un sistema di governance e gestione del rischio coerente e quanto più onnicomprensivo possibile**.

Alcune organizzazioni hanno già beneficiato di un **approccio olistico e integrato** a seguito dell'avvento del GDPR: in quell'occasione, c'è infatti chi ha colto l'opportunità di allineare i propri processi interni in funzione dei requisiti normativi così da gestire in modo più efficace i delicati rischi connessi al trattamento dei dati personali.

L'approccio olistico consente di raggiungere la **conformità alle norme** sul piano non solo formale, ma anche sostanziale, poiché **arricchisce l'identità e la cultura d'impresa** e aiuta a far leva anche sugli obblighi regolatori per conseguire gli **obiettivi strategici e tattici** dell'organizzazione.

Questa è la direzione presa sia dal National Institute of Standards and Technology (NIST) nordamericano con il suo AI RMF, presentato poco sopra, e sia dalla recente **norma ISO/IEC 42001:2023**<sup>1</sup>, il primo stan-

---

<sup>1</sup> Per la sua specificità e potenziale impatto, la norma ISO

dard per i sistemi di gestione dell'IA nelle organizzazioni.

Chi dunque intenda progettare, sviluppare, distribuire o altrimenti implementare l'IA, e intenda farlo in modo responsabile e sostenibile, non può prescindere da un'**analisi complessiva della realtà** in cui l'IA andrà ad operare, dal **coinvolgimento degli stakeholder** interni ed esterni e dalla **definizione di chiare regole** del gioco. Il tutto tenendo a mente che queste tecnologie **evolvono velocemente** e con loro il **quadro dei rischi e dei potenziali impatti**. È una sfida, questa, che vale la pena accettare.

---

sarà oggetto di futura trattazione.

## POSTFAZIONE

*“Il fatto di istruire un selvaggio nell’uso di tecniche avanzate non lo trasforma in una persona civilizzata, ma ne fa solo un selvaggio efficiente [...] Il progresso etico deve accompagnarsi allo sviluppo tecnico ed economico. Mentre insegniamo le tecniche, dobbiamo insegnare anche il rispetto per la dignità e il valore e il carattere sacro della personalità umana. Se non vogliamo che la fine sia peggiore dell’inizio è necessario intraprendere un’azione urgente”*  
(Carlo Cipolla, *Uomini, tecniche, economie*, 1966)

La popolarità e soprattutto la facile accessibilità degli strumenti di intelligenza artificiale (IA) generativa hanno innescato, in poco più di un anno dalla loro apparizione, un progressivo impatto trasformativo sulle attività umane. In un mondo che attraversava discontinuità economiche e geopolitiche, con sussulti causati anche da post pandemie e guerre in corso, è apparsa questa nuova variabile tecnologica portatrice di una potenza trasformativa che ha alimentato con nuove turbative gli scenari socioeconomici globali. Una variabile da gestire con tempestività visto le potenzialità e i rischi di cui è portatrice. E difatti per arrivare quanto prima al rilascio di regolamentazioni e standard sull’uso di questi nuovi strumenti, nel corso del 2023 varie organizzazioni private, l’Europa e i governi nazionali, hanno accelerato le loro attività normative sul tema intelligenza artificiale, sino ad arrivare nella primavera del 2024 all’approvazione del regolamento europeo

sull'intelligenza artificiale da parte del Parlamento Europeo. Sebbene il discorso etico e l'attenzione alle discriminazioni della persona siano elementi cardine delle discussioni in corso in vari tavoli istituzionali e non, è già visibile all'orizzonte il rischio di una frammentazione in apparati normativi con connotati geopolitici diversi che potranno essere utilizzati come elementi di protezione e soprattutto di competitività e di ulteriore "colonizzazione digitale". In parte, i vari approcci alla valutazione dell'impatto dei sistemi di intelligenza artificiale, che sono stati illustrati nei capitoli precedenti, pur avendo molti elementi in comune, sono rappresentazione della frammentazione geopolitica ed istituzionale in corso.

Basterà quindi un regolamento, o più regolamenti, a far trovare l'equilibrio tra le promesse di aumenti di produttività, redditività, maggiore efficienza, potere, ricchezza, controllo, salute e benessere che questi strumenti sembrano realizzare, e allo stesso tempo, garantire la libertà e il governo delle decisioni delle persone?

Per ora vi è un'unica convinzione, suffragata da una pluralità di evidenze empiriche: la società civile e le organizzazioni e, in particolare, le imprese private, saranno i luoghi e i contesti sociali nei quali queste soluzioni tecnologiche sempre più avanzate e dotate di "intelligenza artificiale" avranno il maggiore impatto, situazione che si sta già verificando. Tali effetti, che potrebbero anche essere indiretti e manifestarsi solo nel medio e lungo periodo, rischiano di cambiare il modo stesso di intendere la nostra società, l'attività di impresa e il lavoro delle persone. Prenderne consapevolezza

rappresenta per le persone e le organizzazioni il primo passo di un percorso sull'assunzione di responsabilità sociali della trasformazione digitale e dei rischi d'uso e progettazione dei sistemi di IA e allo stesso tempo significa creare una organizzazione resiliente e responsabile.

Perché per quanto un regolamento nazionale o internazionale potrà dare indicazione sul come valutare le implicazioni etiche, di responsabilità sociale ed ambientale nell'uso e progettazione di queste tecnologie di "intelligenza artificiale", rimane aperta una premessa a tutto ciò, premessa a cui è necessario dare massima attenzione: la consapevolezza delle persone del cambiamento irreversibile che è in atto.

Ciò significa coinvolgere le persone, e le loro organizzazioni, nel prendere coscienza dei fenomeni in atto, averne conoscenza e quindi comprenderne la potenza trasformativa che sta agendo a vari livelli sociali, culturali, economici e politici.

E dato che la sostenibilità è la *"condizione di un modello di sviluppo in grado di assicurare il soddisfacimento dei bisogni della generazione presente senza compromettere la possibilità delle generazioni future di realizzare i propri"*, spetta solo a noi l'averne coscienza degli impatti delle nostre scelte sugli altri, sulle generazioni future, sull'evitare che la persona da soggetto diventi il mero oggetto di applicazioni, procedure o quant'altro.

Ed eccoci al punto focale, è richiesto a tutti noi di sviluppare l'abilità di essere responsabili, di contribuire in modo attivo e partecipativo, al raggiungimento di un progresso tecnologico che, per quanto inesora-

bile, sia intrinsecamente etico e soprattutto mantenga la libertà e la dignità della persona e l'autonomia delle sue scelte. Non ci resta che vivere questa fase cruciale della trasformazione industriale, in atto da secoli, che oggi tocca ambiti sociali, istituzionali, lavorativi e professionali impensabili sino a qualche anno fa, essendo consapevoli dei cambiamenti in atto, attuando azioni etiche concrete e trasparenti.

Quindi le valutazioni d'impatto raccolte ed elencate in questo breve volume, vogliono essere uno stimolo per tutti noi per prendere coscienza della situazione attuale, comprendere quali sono i potenziali impatti trasformativi da conoscere, comprendere e governare.

Ulteriore stimolo è il contributo che segue, "Oltre una visione tecnocentrica", che inquadra su tematiche organizzative ciò di cui tenere conto nel rapporto tra organizzazioni e tecnologie.

*"Mentre insegniamo le tecniche, dobbiamo insegnare anche il rispetto per la dignità e il valore e il carattere sacro della personalità umana. Se non vogliamo che la fine sia peggiore dell'inizio è necessario intraprendere un'azione urgente"* (Carlo Cipolla). L'azione oggi più urgente è quella di prendere coscienza delle forme, talvolta inaspettate, che il presente sta assumendo. È fondamentale, e rappresenta una leva competitiva notevole, studiare e monitorare i fenomeni di rivoluzione digitale e industriale così come capire se, come e quanto gli strumenti di intelligenza artificiale impatteranno sulla vita e sul lavoro delle persone.

*Massimo Manzari*  
CEO - cofounder ReD OPEN

## AI DIGITAL SONAR

A fronte di alcune evidenti e chiare **opportunità** legata alla crescita e alla diffusione dell'AI, non mancano anche aree di **rischio** che richiedono di adottare un approccio "responsabile". Che cosa significa adottare un approccio responsabile di fronte all'introduzione di nuove tecnologie lo abbiamo scoperto con l'avvento e le successive riflessioni legate all'introduzione del GDPR, che ha posto l'accento sull'importanza di saper gestire gli equilibri le spinte all'utilizzo vasto e creativo dei dati disponibili e l'esigenza di proteggere, tra questi, quelli sensibili e personali. La risposta all'indeterminatezza su dove tracciare i confini tra uso e abuso ha portato all'introduzione dell'approccio "**by design**": un modo per suggerire alle imprese, in assenza di leve chiare ed esperienze pratiche precise, di ... "pensarci prima", seriamente e responsabilmente. In ragionevole continuità, sia normativa che di buon senso, l'approccio *by design* si presta anche allo sviluppo di una gestione responsabile dell'AI, per sfruttarne appieno le opportunità anticipando i rischi, non solo di data protection, che potrebbero manifestarsi. Come fare?

Per provare a dare una risposta a questa domanda, nasce **AI DIGITAL SONAR**, un'iniziativa concepita da ReD OPEN con il sostegno di EUMETRA, il cui obiettivo è quello di contribuire ad avviare proattivamente un processo di consapevolezza diffusa tra coloro che prendono decisioni, utilizzano, applicano, proget-

tano e comunicano l'Intelligenza Artificiale. L'ambizione di AI DIGITAL SONAR è quella di favorire un **approccio responsabile “by design”** all'innovazione tecnologica, per massimizzarne i vantaggi e contenere, ridurre, evidenziare e magari evitare i potenziali rischi per persone, ambiente e società.

Per promuovere un approccio responsabile all'innovazione tecnologica e in particolare per porre attenzione al ruolo dell'Intelligenza Artificiale nei percorsi di adozione e trasformazione digitale, AI DIGITAL SONAR ritiene indispensabile fare **attività di ricerca e “ascolto”**, stimulate da un'analisi esplorativa. Questa analisi, condotta da EUMETRA, ha l'obiettivo di cogliere la **“percezione dei cittadini”** e prevede un servizio di **monitoraggio** della “reputazione responsabile” che coinvolge un ampio campione di imprese che operano nel nostro Paese.

L'obiettivo esplicito di AI DIGITAL SONAR è quello di **condividere pratiche e riflessioni** per identificare, tra percorsi e contesti diversi, eventi e momenti di confronto organizzati in collaborazione con il Dipartimento di Giurisprudenza dell'Università di Milano-Bicocca, un **approccio responsabile** alla gestione by design per la protezione dei dati e per l'adozione di tecnologie di AI.

Per maggiori informazioni su AI DIGITAL SONAR:  
**aidigitalsonar@redopen.it**





## RIFERIMENTI BIBLIOGRAFICI

- Bijker, W.E. (1999), *Of bicycles, bakelites and bulbs. Towards a theory of sociotechnical change*, MIT Press, Cambridge.
- Bijker, W.E., Hughes, T.P., Pinch, (1987) (a cura di), *The Social Construction of Technological Systems. New Directions in the Sociology and History of Technology*, MIT Press, Cambridge.
- Catino, M. (2006), *Da Chernobyl a Linate. Incidenti tecnologici o errori organizzativi?*, Mondadori, Milano.
- Cicourel, A.V. (1996), *Ecological validity and white rooms effects*, in «Pragmatics & Cognition», 4 (2): 221-264.
- Cipolla, C.M. (1966), *Uomini tecniche economie*, Feltrinelli, Milano.
- Commissione Europea (2018), *L'intelligenza artificiale per l'Europa*, COM (2018) 237 final.
- Commissione Europea (2021), *Proposta di Regolamento del Parlamento Europeo e del Consiglio che stabilisce regole armonizzate sull'Intelligenza Artificiale (Legge sull'Intelligenza Artificiale) e modifica alcuni atti legislativi dell'Unione*, COM (2021) 206 final.
- Consiglio dell'Unione Europea (2024), *Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts - Analysis of the final compromise text with a view to agreement*.
- Corradi, G., Gherardi, S., Verzelloni, L. (2010), *Through the practice lens: Where is the bandwagon of practice-*

- based studies heading?, in «Management Learning», 41, 3, 265-83.
- Galbraith, J.K. (1967), *The New Industrial State*, Houghton Mifflin, Boston.
- Gherardi, S. (2009) (a cura di), *Apprendimento tecnologico e tecnologie di apprendimento*, Il Mulino, Bologna.
- Godin, B. (2015), *Innovation contested: The idea of innovation over the centuries*, Routledge, London.
- Gruppo indipendente di esperti ad alto livello sull'intelligenza artificiale (High-Level Expert Group on AI – AI HLEG) (2019), *Linee guida etiche per l'IA basata sulla fiducia* (<https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>).
- Gruppo indipendente di esperti ad alto livello sull'intelligenza artificiale (High-Level Expert Group on AI – AI HLEG) (2020), *Assessment List for Trustworthy Artificial Intelligence for self-assessment* (<https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment>).
- Kimberly J.R., Evanisko M.J. (1981), *Organizational innovation*, in «Academy of Management Journal», 24, 4, 689-713.
- Latour, B. (1988), *The pasteurization of France*, Harvard University Press, Cambridge.
- Lepore, J. (2014), *The disruption machine*, in «The New Yorker», 23 giugno.
- Leslie, D. (2019), *Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector*, The Alan Turing Institute.
- Norman, D.A. (1993), *Things that make us smart. Defending*

- Human Attributes in the Age of the Machine, Addison-Wesley, Reading.
- Norman, D.A. (1998), *The invisible technology*, MIT Press, Cambridge.
- Parlamento Europeo (2024), Comunicato stampa: Il Parlamento europeo approva la legge sull'intelligenza artificiale, <https://www.europarl.europa.eu/news/it/press-room/20240308IPR19015/il-parlamento-europeo-approva-la-legge-sull-intelligenza-artificiale>.
- Piana, D., Verzelloni, L. (2022). Epistemic communities meet communities of practices, in «Rivista di Digital Politics», II (1-2), 221-246.
- Postman, N. (2000), *Building a Bridge to the 18th Century*, Vintage Books, New York.
- Suchman, L.A. (1987), *Plans and situated actions: The problem of human-machine communication*, Cambridge University Press, New York.
- UNESCO (2021), *Recommendation on the Ethics of Artificial Intelligence*.
- Van de Ven, A.H. (1986), Central problems in the management of innovation, in «Management Science», 32, 5, 590-607.
- Verzelloni, L. (2019), *Pratiche di sapere*, Rubbettino, Soveria Mannelli.
- Verzelloni, L. (2020a), *Paradossi dell'innovazione*, Carocci, Roma.
- Verzelloni, L. (2020b), La vana ricerca della neutralità: razionalità ibride e bilanciamento di poteri, in tempi di crisi e non, in «Federalismi.it», 28, 197-211.
- World Economic Forum (2022), *Unpacking AI Procurement in a Box: Insights from Implementation*.

## ELENCO DELLE VALUTAZIONI DELL'IMPATTO ALGORITMICO DESCRITTE

In ordine di apparizione nel testo:

- [1] Government of Canada (2019), Algorithmic Impact Assessment Tool.
- [2] Federal CIO Council (United States) (2021), Algorithmic Impact Assessment Tool.
- [3] Digital Dubai Authority (2020). AI systems ethics self-assessment.
- [4] Krafft, T., Hauer, M., & Fetic, L. (2020), From principles to practice: an interdisciplinary framework to operationalise AI ethics. VDE and Bertelsmann Stiftung.
- [5] Various Authors (2018), Artificial intelligence impact assessment. Technical report, Platform voor InformatieSamenleving (ECP).
- [6] World Economic Forum (2020), Unlocking Public Sector AI. AI Procurement in a Box: Workbook.
- [7] Andrade, N. N. G. D., & Kontschieder, V. (2021), AI Impact Assessment: A Policy Prototyping Experiment.
- [8] Michele Loi, Anna Mätzener, Angela Müller, & Matthias Spielkamp (2021), Automated Decision-Making Systems in the Public Sector. An Impact Assessment Tool for Public Authorities. Algorithm Watch.
- [9] Ada Lovelace Institute. Algorithmic impact assessment: a case study in healthcare, February 2022.
- [10] Floridi, L., Holweg, M., Taddeo, M., Amaya Silva, J., Mökander, J., & Wen, Y. (2022), capAI - A Procedure for Conducting Conformity Assessment of AI Systems in Line with the EU Artificial Intelligence Act.

- [11] Netherlands Court of Audit (2021), Audit Framework for Algorithms.
- [12] European Law Institute (2022), Model rules on impact assessment of automated decision-making systems used by public administration.
- [13] Netherlands Court of Audit (2022), Fundamental rights and algorithm impact assessment (FRAIA).
- [14] UNESCO (2023), Ethical Impact Assessment. A Tool of the Recommendation on the Ethics of Artificial Intelligence.
- [15] Netherlands Ministry of Infrastructure and Water Management (2023), AI Impact Assessment. A tool to set up responsible AI projects.
- [16] Governo di Singapore (2024), Proposed Model AI Governance Framework For Generative AI.
- [17] National Institute of standards and technology (2023), Artificial Intelligence Risk Management Framework (AI RMF).

## CONTESTO

Negli ultimi decenni, l'innovazione, soprattutto tecnologica, è diventata la "stella polare" che orienta governi, istituzioni internazionali, imprese private, comunità scientifica e società civile. Eppure, **anche l'innovazione ha un "lato oscuro"** che, se non adeguatamente governato, **può generare degli effetti perversi e paradossali, che possono manifestarsi a livello individuale, organizzativo e sistemico.**

## OGGI, PIÙ CHE MAI, DOBBIAMO PORCI UNA NUOVA DOMANDA:

**come trovare un equilibrio tra le promesse dell'intelligenza artificiale, in termini di incremento di produttività, redditività ed efficienza delle imprese e l'esigenza di garantire, al contempo, sostenibilità e coerenza organizzativa,** rendendo consapevoli le persone dei cambiamenti in atto?

## NOI TI AIUTIAMO A RISPONDERE CON AI CHECK & GO™



Accompagna le imprese in percorsi di responsabilità e resilienza, e prepara l'organizzazione nell'attuare al proprio interno regolamenti quali EU AI ACT piuttosto che norme come ISO42001.



Ambisce ad attivare un processo di consapevolezza diffusa in coloro che decidono, utilizzano, progettano e comunicano l'intelligenza artificiale intesa come strumento d'innovazione sociale.



Genera fiducia e attesta la tua trasparenza sia all'interno che all'esterno della tua organizzazione.

## A CHI CI RIVOLGIAMO?

1. I destinatari dello strumento sono, **in primo luogo, gli amministratori d'impresa e i C-Level** che, attraverso le loro decisioni, influenzano la strategia aziendale.



2. La platea di destinatari comprende anche **coloro che, a vario titolo, utilizzano questi strumenti all'interno e all'esterno delle organizzazioni**, non di rado senza avere gli strumenti per comprenderne e valutarne in modo adeguato gli impatti sociali e organizzativi.



## COS'È E COME FUNZIONA AI CHECK & GO™?

**AI CHECK & GO™ è un questionario di auto-valutazione** della consapevolezza dell'organizzazione rispetto agli impatti e ai rischi che derivano dall'uso dell'Intelligenza Artificiale.

Questo strumento **è il punto di partenza per chi vuole abbracciare un approccio positivo all'Intelligenza Artificiale.**

**Il questionario viene somministrato individualmente a un campione di dipendenti** afferenti a diversi settori, **da concordare con l'organizzazione**, in funzione della sua grandezza, del grado di strutturazione aziendale e del ruolo dell'IA nella value chain (provider, deployer, ecc.).



Il questionario si articola in **120 domande**, che richiedono un tempo medio di risposta di pochi secondi, **organizzate in 7 macro-sezioni**:



**1. CONTESTO GENERALE**



**5. NON DISCRIMINAZIONE E EQUITÀ**



**2. SCOPI E OBIETTIVI DICHIARATI**



**6. PRIVACY, DATA PROTECTION E DATA GOVERNANCE**



**3. IMPATTO ORGANIZZATIVO**



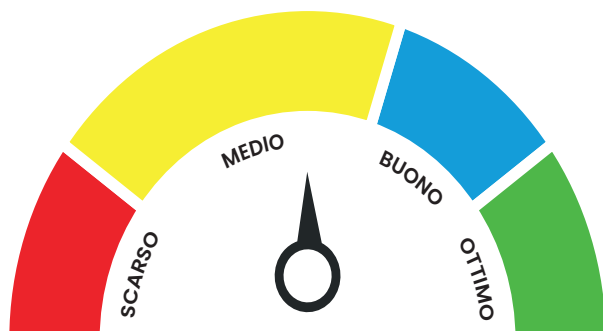
**4. TRASPARENZA E ACCOUNTABILITY**



**7. IMPATTO AMBIENTALE**

## RISULTATI

Allo scopo di **restituire dei risultati di facile comprensione sulle possibili aree di rischio**, **AI CHECK & GO™** utilizza dei **tachimetri** (*speedometer*) su una scala da 100 punti (**grado di consapevolezza** scarso, medio, buono e ottimo). All'impresa che adotta questo strumento viene restituito un rapporto di ricerca realizzato ad hoc (*tailor-made*), comprensivo di alcune proposte per migliorare consapevolezza e resilienza aziendale.





# AI CHECK & GO™

*Un approccio positivo all'Intelligenza Artificiale*

A fronte di un risultato ottimo o buono dell'autovalutazione **AI CHECK & GO™**, ReD OPEN rilascia l'attestazione **AI ResponsAbility by Design™**.

Otterranno l'attestazione le imprese che si stanno impegnando in un percorso di miglioramento continuo sui temi dell'impatto etico-sociale-normativo delle tecniche di IA e che si configurano come un provider responsabile per i propri clienti business e/o consumatori finali. Per cogliere i processi di trasformazione all'interno dei contesti aziendali, l'attestazione viene rinnovata su base annuale.



ReD OPEN è uno spin-off dell'Università degli Studi Milano Bicocca iscritto al registro delle start-up innovative che attraverso strumenti quali AI Check & Go™, e metodi mutuati dalla ricerca, sostiene sin dalla sua nascita nel 2019 la diffusione di un approccio consapevole, responsabile e sostenibile alla trasformazione digitale, alla progettazione e all'utilizzo di strumenti di Intelligenza Artificiale.

 **REDOPEN**  
*toward a conscious present.*





ReD OPEN è uno spin-off dell'**Università di Milano-Bicocca**, nato con l'intento di accompagnare e aiutare le imprese per affrontare percorsi di innovazione, transizione digitale e ricorso all'Intelligenza Artificiale in modo «responsible by design». Con competenze ed esperienze multidisciplinari, ReD OPEN propone percorsi di accompagnamento e di riconfigurazione dei modelli organizzativi e di business, in chiave responsabile.

Il ReD OPEN Competence Center:

- ha la missione di **diffondere nelle imprese la ricerca** non ancora applicata e di rappresentare la forma permanente di un modello di business orientato al trasferimento tecnologico tra ricerca e impresa;
- media, per conto delle imprese, **linguaggi e processi potenzialmente innovativi** ma ancora lontani dall'essere “mainstream”;
- si fonda su **logiche partecipative e condivise**, atte a trasferire alle aziende modelli e piattaforme operative di governance dell'innovazione che abilitano la trasformazione digitale in modo «consapevole».

Per approfondimenti:

[www.redopen.it](http://www.redopen.it)



A distanza di due anni dalla pubblicazione di *Governare l'Intelligenza Artificiale* – e a seguire l'approvazione del regolamento europeo AI ACT – con questo volume rendiamo disponibile una panoramica non esaustiva di alcuni strumenti di valutazione dell'impatto algoritmico predisposti finora da organizzazioni diverse, governative e private.

Lo scopo di questa pubblicazione è offrire al lettore alcune semplici indicazioni per orientarsi in un paesaggio normativo complesso che sarà presto lo standard per operare sul mercato, invitando a porre l'accento su sistemi di governance e gestione del rischio sempre più integrati.

Il volume si inserisce come primo contributo pubblico del progetto AI Digital Sonar, il cui obiettivo è quello di contribuire ad avviare proattivamente un processo di consapevolezza diffusa tra le persone, e soprattutto tra coloro che prendono decisioni, utilizzano, applicano, e progettano tecnologie di Intelligenza Artificiale.

[www.ledizioni.it](http://www.ledizioni.it)