

**M. Calvano, A. Curci, A. Esposito,
R. Lanzilotti, A. Piccinno**

L'AI Act come rampa per un'intelligenza artificiale simbiotica

Abstract: The AI Act as a Road Toward Symbiotic Artificial Intelligence. Artificial Intelligence (AI) is rapidly spreading, impacting many areas of our lives. In this context, to ensure the protection of the fundamental rights of human beings and of the environment that surrounds them, it is necessary to create systems that comply with the law. In this regard, the European Union has introduced the AI Act, a regulatory framework aimed at governing the development and use of AI through a risk-based approach. At the same time, researchers emphasize the need to design human-centered AI systems capable of fostering a symbiotic relationship between humans and AI. The objective of Symbiotic AI systems is to enhance rather than replace humans through continuous collaboration and learning. Considering the need to create legally compliant Symbiotic AI systems, this chapter describes the FrameSAI framework (*Framework to create Symbiotic Artificial Intelligence*), which aims to guide practitioners throughout the creation process of Symbiotic AI systems compliant with the AI Act.

Keywords: Symbiotic AI, European AI Act, Symbiosis-by-design, Human-Centered.

SOMMARIO: 1. Introduzione – 2. Intelligenza artificiale simbiotica e centrata sulle persone – 3. L'AI Act per la simbiosi – 4. Indicazioni pratiche – 5. Conclusioni.

1. Introduzione

L'intelligenza artificiale (generalmente abbreviata in IA o, partendo dall'inglese *artificial intelligence*, AI) è ormai ubiqua. Ci interfacciamo con sistemi di IA giornalmente: basta pensare ai motori di raccomandazione che permettono a servizi di *streaming* digitale (per esempio, Spotify e Netflix) di suggerire il prossimo contenuto da visualizzare. Un altro esempio è fornito dai *chatbot*, a partire da quelli con cui siamo costretti a interfacciarci quando vogliamo contattare l'assistenza clienti di varie aziende, fino a quelli basati su IA generativa e *large language*

models (LLM), come ChatGPT, che utilizziamo per ottenere informazioni o per scrivere testo.

In aggiunta ai diversi domini sopra citati, i sistemi di IA promettono di rivoluzionare settori che impattano in modo significativo la vita degli esseri umani, come quello medico e legale. Per citare un esempio, vari studi mirano a realizzare sistemi di IA in grado di riconoscere e diagnosticare malattie come il morbo di Alzheimer o tumori cerebrali nelle loro fasi iniziali¹. Un ulteriore esempio è fornito dai casi in cui l'IA è utilizzata in contesti giudiziari per riconoscere sospettati di un crimine e intervenire in maniera tempestiva².

Tuttavia, l'uso di tecniche di IA non è privo di rischi, ed è una scelta che va ponderata adeguatamente. Infatti, le tecniche di IA attualmente maggiormente utilizzate (in quanto più performanti) sono tecniche di *machine learning*, o più precisamente di *deep learning*, che sono spesso caratterizzate sia dalla loro impossibilità di rendere note le motivazioni dietro le decisioni prese, sia dall'impossibilità di lasciare il controllo delle azioni all'essere umano. È noto, infatti, quanto sia pericoloso automatizzare un processo senza considerare appieno l'elemento umano con cui, inevitabilmente, il sistema deve interfacciarsi: i suoi utenti. Esistono diversi esempi a supporto di questa necessità, raccolti sul sito web *AI Incident Database*³, il quale riporta dati riguardo agli incidenti causati dall'intelligenza artificiale.

¹ F. Altay *et al.*, "Preclinical stage Alzheimer's disease detection using magnetic resonance image scans", in *Proceedings of the AAAI conference on artificial intelligence*, 35 (2021), pp. 15088-15097, <https://doi.org/10.1609/aaai.v35i17.17772>; G. Castellano *et al.*, "Automated Detection of Alzheimer's Disease: A Multi-Modal Approach with 3D MRI and Amyloid PET", in *Scientific Reports* 14 (2024), n. 5210, <https://doi.org/10.1038/s41598-024-56001-9>; A. Curci, A. Esposito, "Detecting Brain Tumors Through Multimodal Neural Networks", in *Proceedings of the 13th International Conference on Pattern Recognition Applications and Methods*, SCITEPRESS - Science and Technology Publications, 26 febbraio 2024, pp. 995-1000, <https://doi.org/10.5220/0012608600003654>.

² S. Raaijmakers, "Artificial Intelligence for Law Enforcement: Challenges and Opportunities", in *IEEE Security & Privacy* 17 (2019), n. 7, pp. 74-77, <https://doi.org/10.1109/MSEC.2019.2925649>.

³ <https://incidentdatabase.ai>

In tale contesto, funge da ammonimento il rapporto della *National Transportation Safety Board* (NTSB, un'agenzia investigativa indipendente del Governo degli USA che indaga ed emette rapporti in merito agli incidenti che coinvolgono aeroplani, navi, treni, oleodotti e gasdotti) su un incidente tra una Tesla e un autocarro con rimorchio. Il passaggio più interessante è il seguente (tradotto dagli autori):

Poiché oltre il 90% degli incidenti stradali è attribuito a errori umani, l'automazione nei veicoli offre un notevole potenziale per salvare decine di migliaia di vite ogni anno, sostituendo progressivamente il conducente. Tuttavia, introdurre l'automazione in un ambiente così complesso e non strutturato sarà molto impegnativo e dovrà essere affrontato con attenzione e grande cautela.

La storia iniziale dell'automazione in aviazione comprende molti esempi in cui l'automazione è stata introdotta semplicemente perché l'industria possedeva la capacità tecnologica per farlo. L'esperienza ha insegnato che automatizzare “perché possiamo” non significa necessariamente migliorare il funzionamento del sistema uomo-automazione. Questo ha portato a un'evoluzione verso un'automazione “umano-centrica”, il cui obiettivo è migliorare le prestazioni complessive del sistema uomo-automazione. Questo incidente è un esempio di ciò che può accadere quando l'automazione viene introdotta “perché possiamo” senza un'adeguata considerazione del fattore umano.⁴

In risposta ai rischi associati all'uso dell'IA, l'Unione Europea (UE) ha introdotto il *framework* legislativo comunemente denominato “AI Act”⁵. L'AI Act è caratterizzato da un insieme di limitazioni e indicazioni sull'uso dell'IA in base al livello di *rischio* associato in contesti specifici. Più precisamente, l'AI Act

⁴ National Transportation Safety Board, *Collision Between a Car Operating with Automated Vehicle Control Systems and a Tractor-Semitrailer Truck Near Williston, Florida, May 7, 2016*, Highway Accident Report NTSB/HAR-17/02 (National Transportation Safety Board, 2017), <https://www.ntsb.gov/investigations/accidentreports/reports/har1702.pdf>.

⁵ Regulation of the European Parliament and of the Council Laying down Harmonised Rules on Artificial Intelligence and Amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (2024).

individua quattro livelli di rischio: minimale, limitato, alto, inaccettabile; maggiore è il livello di rischio, più restrizioni ci sono sull'uso dell'IA – fino ad arrivare al totale divieto in contesti con rischio inaccettabile (per esempio, per applicazioni di *social scoring*).

Parallelamente all'introduzione nel panorama legislativo dell'AI Act, la comunità accademica ha riconosciuto i rischi dell'uso di IA, portando alla nascita di una nuova corrente di ricerca denominata “intelligenza artificiale centrata sulle persone” (o, in inglese, *human-centered AI*, abbreviato in HCAI)⁶. Differentemente dal classico approccio all'IA, che si focalizza sulle *performance* tecniche del modello intelligente, la HCAI mira a creare sistemi che siano comprensibili, controllabili e allineati con gli obiettivi degli esseri umani – ossia noi, gli utenti finali degli stessi sistemi. La HCAI cambia il modo in cui gli utenti sono concepiti all'interno del ciclo di vita di un sistema di IA: la visione classica li pone a semplici osservatori passivi, che si limitano a ricevere gli *output* del sistema; al contrario, la HCAI li eleva a una posizione attiva, rendendo il processo decisionale un processo collaborativo. Questa visione dell'utente è ulteriormente specializzata nel concetto di IA *simbiotica*⁷.

Prendendo in prestito dalla terminologia comunemente usata in biologia, l'IA simbiotica instaura un rapporto molto più profondo con i propri utenti umani. Questo rapporto, collaborativo per natura, è caratterizzato da un mutuale miglioramento delle due forme di intelligenza (umana e artificiale). Infatti, un'IA simbiotica mira a *migliorare* (invece di sostituire) il proprio utente umano, aumentandone le capacità e le prestazioni; al contempo, l'IA viene costantemente migliorata apprendendo continuamente dalle azioni che compie l'utente.

Benché l'approccio basato sul rischio dell'AI Act e le visioni di IA simbiotica e centrata sull'essere umano possano sembrare

⁶ B. Shneiderman, *Human-Centered AI*, 1^a ed., Oxford University Press, Oxford, 2022, <https://doi.org/10.1093/oso/9780192845290.001.0001>.

⁷ S.S. Grigsby, “Artificial Intelligence for Advanced Human-Machine Symbiosis”, in D.D. Schmorow, C.M. Fidopiastis (eds.), *Augmented Cognition: Intelligent Technologies*, vol. 10915, Lecture Notes in Computer Science, Springer International Publishing, Cham, 2018, https://doi.org/10.1007/978-3-319-91470-1_22.

distanti fra loro, sono in realtà estremamente vicine. Infatti, uno dei principi fondamentali dell'AI Act è la necessità di sviluppare sistemi “*human-centric*” che lascino il controllo all'essere umano, rispettandone i diritti e le norme etiche della società in cui viviamo. Tuttavia, una delle principali difficoltà dell'implementazione dell'AI Act è proprio nella differenza di linguaggio tra la legislazione e la matematica (linguaggio basilare dell'IA). Questo capitolo si pone quindi l'obiettivo di delineare un “ponte” fra i due mondi, fornendo indicazioni concrete. Saranno descritti linee guida e criteri di successo, estratti dalla letteratura e componenti di *FrameSAI* (dall'inglese: *Framework to create Symbiotic Artificial Intelligence*)⁸, per la creazione di sistemi di IA simbiotici nel pieno rispetto dell'AI Act – punto di partenza della definizione delle linee guida e dei criteri di successo stessi.

2. Intelligenza artificiale simbiotica e centrata sulle persone

In questa sezione, saranno forniti il contesto e la terminologia necessari a guidare la lettura di *FrameSAI*. Il fulcro delle linee guida e dei criteri di successo presentati in questo capitolo è il concetto di intelligenza artificiale simbiotica, che a sua volta, si basa sulla più generale disciplina dell'intelligenza artificiale centrata sulle persone (si veda fig. 1).

⁸ M. Calvano *et al.*, “FrameSAI: a Three-Layer Framework to Create Symbiotic AI Systems”, in *Workshop “AI Responsabile e Affidabile” co-located with the 5th National Conference on Artificial Intelligence, organized by CINI (Ital-IA 2025)*, CEUR Workshop Proceedings, 2025.

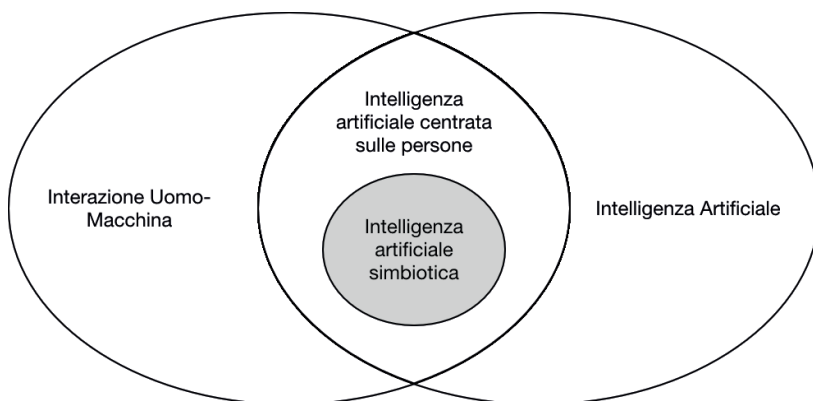


Figura 1. Relazione tra IA simbiotica e IA centrata sulle persone

L'interazione persona-macchina è un campo multidisciplinare (è basato sull'informatica, ma anche su psicologia, ergonomia, ecc.) che si focalizza sulla progettazione e sulla valutazione di sistemi interattivi (tipicamente informatici) enfatizzando l'importanza dell'interazione tra gli utenti e i sistemi⁹. L'intelligenza artificiale, invece, si occupa dello studio di agenti che mostrano caratteristiche assimilabili a un agente "intelligente", essendo quindi in grado di percepire il mondo loro circostante e agire autonomamente¹⁰. L'intelligenza artificiale centrata sulle persone (dall'inglese *human-centered artificial intelligence*, HCAI), quindi, ha l'obiettivo di progettare e valutare sistemi interattivi *intelligenti*, focalizzandosi sull'interazione tra essi e i loro utenti, affinché siano affidabili, sicuri e degni di fiducia¹¹. Più precisamente, i sistemi di intelligenza artificiale centrati sulle persone sono definiti come segue:

I sistemi di intelligenza artificiale centrati sulle persone (Human-Centered AI) sono sistemi di IA progettati, sviluppati e valutati coinvolgendo gli utenti nel processo, con l'obiettivo di aumentare le prestazioni e la soddisfazione delle persone in specifici contesti. I sistemi HCAI mirano a esse-

⁹ A. Dix (ed.), *Human-Computer Interaction*, 3^a ed., Pearson Prentice-Hall, Upper Saddle River, 2010.

¹⁰ S.J. Russell, P. Norvig, *Artificial Intelligence: A Modern Approach*, 3^a ed., Pearson, London, 2016.

¹¹ B. Shneiderman, *Human-Centered AI*, cit.

re utili e usabili per specifici utenti, i quali possono essere descritti attraverso un modello formale, al fine di raggiungere i propri obiettivi all'interno del loro contesto d'uso, garantendo al contempo affidabilità, sicurezza d'uso e fiducia¹².

Uno dei capisaldi dell'IA centrata sulle persone è, quindi, la necessità di sviluppare sistemi che *potenzino* le capacità umane, preservino l'autonomia dell'utente e includano meccanismi di responsabilità e *governance*¹³. Partendo da queste premesse, è emersa una nuova direzione identificata con il termine "IA *simbiotica*"¹⁴ che immagina sistemi progettati non solo per potenziare le capacità umane in un dato momento, ma anche per apprendere ed evolversi continuamente attraverso l'interazione.

L'IA simbiotica aspira a promuovere una relazione in cui sia il sistema di IA sia l'utente umano crescono e si adattano *insieme*, migliorando reciprocamente le proprie capacità in una collaborazione continua e reciprocamente vantaggiosa tra intelligenza artificiale e intelligenza umana. Così come avviene per la HCAI, quando correttamente implementati, i sistemi simbiotici mirano a supportare i propri utenti senza sostituirli, riconoscendo che gli utenti restano partecipanti essenziali nel ciclo d'interazione: invece di automatizzare completamente decisioni o azioni, i sistemi di IA possono apprendere dai comportamenti, dalle preferenze e dalle correzioni degli utenti, adattandosi nel tempo per allinearsi sempre meglio agli obiettivi umani. Possiamo quindi definire un sistema di IA simbiotica come segue:

I sistemi di Intelligenza Artificiale Simbiotica (*Symbiotic AI*, SAI) sono sistemi HCAI basati su una collaborazione continua e profonda tra esseri umani e intelligenza artificiale, ovvero una simbiosi tra intelligenza umana e intelligenza artificiale. In questi sistemi, esseri umani e IA si potenziano

¹² G. Desolda *et al.*, "From Human-Centered to Symbiotic Artificial Intelligence: A Focus on Medical Applications", in *Multimedia Tools and Applications* 84 (2024), n. 27, pp. 32109-32150, <https://doi.org/10.1007/s11042-024-20414-5>.

¹³ B. Shneiderman, "Bridging the Gap Between Ethics and Practice: Guidelines for Reliable, Safe, and Trustworthy Human-Centered AI Systems", in *ACM Transactions on Interactive Intelligent Systems* 10 (2020), n. 4, pp. 1-31, <https://doi.org/10.1145/3419764>.

¹⁴ S.S. Grigsby, *op. cit.*

reciprocamente, bilanciando i rispettivi punti di forza e di debolezza, senza compromettere né l'autonomia degli esseri umani né le prestazioni dell'IA, mantenendo il controllo delle decisioni del sistema nelle mani dell'essere umano. In un sistema SAI, l'IA trae beneficio da un flusso continuo di nuovi dati forniti dagli utenti per perfezionarsi, mentre gli esseri umani traggono vantaggio dal miglioramento delle prestazioni e delle conoscenze dell'IA¹⁵.

La crescente adozione e integrazione di sistemi di IA in vari processi decisionali a cui stiamo assistendo rendono fondamentale una modifica alle modalità di progettazione, sviluppo e valutazione di sistemi IA. L'uso sempre crescente di IA in domini particolarmente delicati (come la medicina) rende infatti necessario un approccio che consenta agli utenti di mantenere il controllo del processo decisionale, lasciando loro la stessa autonomia di giudizio che avrebbero senza l'uso di sistemi intelligenti, ma supportandoli per rendere le loro attività più rapide e ridurre il numero di errori. È questa necessità che sta alimentando la crescita dell'adozione del "paradigma simbiotico" sopra descritto.

Infatti, i sistemi di IA simbiotica, collaborando con i loro utenti umani invece di sostituirli, garantiscono la tutela dei loro diritti fondamentali, oltre che dell'ambiente in cui sono utilizzati. In questo contesto, l'AI Act non è solo una componente dell'ambiente circostante, né tantomeno una semplice legge che impone delle limitazioni o istituisce dei requisiti (come la necessità di registrare sistemi ad alto rischio in un apposito *database* della Commissione Europea), ma diventa un'utile guida che supporti i progettisti di sistemi intelligenti a creare IA che non rappresentino un rischio per gli esseri umani¹⁶.

3. L'AI Act per la simbiosi

In questo contesto, è necessario guidare i progettisti verso l'adozione di un approccio per la creazione di sistemi di IA

¹⁵ G. Desolda *et al.*, "From Human-Centered to Symbiotic Artificial Intelligence", cit.

¹⁶ M. Wörsdörfer, "Mitigating the Adverse Effects of AI with the European Union's Artificial Intelligence Act: Hype or Hope?", in *Global Business and Organizational Excellence*, 14 novembre 2023, joe.22238, <https://doi.org/10.1002/joe.22238>.

chiamato *symbiosis-by-design*, che mira a creare sistemi di IA simbiotica che rispettano gli standard etici delineati nell'AI Act. Tale approccio stabilisce requisiti specifici per i sistemi di IA progettati per lavorare a stretto contatto con gli esseri umani, che si basano su quattro principi fondamentali, che caratterizzano l'IA simbiotica: Livello di Automazione, Trasparenza, Equità e Protezione¹⁷.

Il principio del *livello di automazione* si traduce in meccanismi di interazione specifici che devono essere stabiliti per raggiungere una relazione simbiotica tra le due parti coinvolte. Ad esempio, i sistemi di IA devono essere progettati garantendo un adeguato controllo e supervisione umana (la scelta o il compromesso tra le due opzioni è fortemente influenzata dal dominio specifico in cui l'IA è utilizzata). Un'attenzione particolare deve essere posta anche alla *trasparenza* dell'IA, in modo che possa essere monitorata dagli esseri umani. Affinché ciò sia possibile, gli esseri umani devono poter comprendere il comportamento dell'IA, che deve fornire spiegazioni chiare che consentono di comprendere le motivazioni dell'*output* fornito. Inoltre, questo principio supporta anche quello dell'*equità*, poiché i sistemi di IA devono evitare comportamenti discriminatori e assicurare l'uso di informazioni corrette e legittime. Infine, durante il processo di interazione, è importante assicurare la protezione degli esseri umani da comportamenti malevoli dell'IA salvaguardando la loro *privacy* (restano infatti validi i requisiti del GDPR¹⁸) e prevenendo qualsiasi forma di danno, sia al singolo utente umano che alla più vasta società e all'ambiente circostante¹⁹.

¹⁷ J. Mökander *et al.*, “Conformity Assessments and Post-Market Monitoring: A Guide to the Role of Auditing in the Proposed European AI Regulation”, in *Minds and Machines*, 32 (2022), n. 2, pp. 241-268, <https://doi.org/10.1007/s11023-021-09577-4>; M. Calvano *et al.*, “Building Symbiotic AI: Reviewing the AI Act for a Human-Centred, Principle-Based Framework”, in *Minds and Machines*, (2025), <https://doi.org/10.48550/arXiv.2501.08046>.

¹⁸ Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 32016R0679 (2016), <http://data.europa.eu/eli/reg/2016/679/oj>.

¹⁹ M. Calvano *et al.*, “Building Symbiotic AI”, cit.

Per garantire la creazione di sistemi di IA simbiotica che siano conformi con l'AI Act, è necessario prendere in considerazione, oltre ai principi, anche tre proprietà fondamentali: Sostenibilità, Affidabilità e Robustezza. È importante specificare che queste proprietà non sono una condizione sufficiente per la creazione di sistemi di IA simbiotica, ma emergono come componenti fondamentali da considerare durante il processo di creazione assieme ai principi. Difatti, la fig. 2 mostra la connessione tra i principi e le proprietà.

Le proprietà identificate sono definite nel seguente modo. La *Sostenibilità* riguarda la necessità di sviluppare sistemi di IA conformi con la legge e con i requisiti di sostenibilità, in grado di essere affidabili e resistenti anche di fronte a cambiamenti ambientali²⁰. L'*Affidabilità* si riferisce alla capacità dei sistemi di IA di consentire agli esseri umani di supervisionare e/o controllare in modo adeguato le loro prestazioni, mantenendo un livello appropriato di automazione, adottando comportamenti equi e rispettando i diritti umani²¹. La *Robustezza*, infine, indica la capacità dei sistemi di operare in modo affidabile ed efficace in diverse condizioni, comprese quelle inaspettate o particolarmente complesse, senza compromettere le proprie funzioni o la sicurezza²².

²⁰ U. Pagallo *et al.*, “The Environmental Challenges of AI in EU Law: Lessons Learned from the Artificial Intelligence Act (AIA) with Its Drawbacks”, in *Transforming Government: People, Process and Policy*, 16 (2022), n. 3, pp. 359-376, <https://doi.org/10.1108/TG-07-2021-0121>.

²¹ G. Stettinger *et al.*, “Trustworthiness Assurance Assessment for High-Risk AI-Based Systems”, *IEEE Access*, 12 (2024), pp. 22718-45, <https://doi.org/10.1109/ACCESS.2024.3364387>; European Commission: Directorate-General for Communications Networks, Content and Technology, *The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for Self Assessment*. (Publications Office, 2020), <https://doi.org/10.2759/002360>.

²² J. Tallberg *et al.*, “AI Regulation in the European Union: Examining Non-State Actor Preferences”, *Business and Politics* 26, fasc. 2 (2024): 218–39, <https://doi.org/10.1017/bap.2023.36>; Johann Laux *et al.*, “Trustworthy Artificial Intelligence and the European Union AI Act: On the Conflation of Trustworthiness and Acceptability of Risk”, *Regulation & Governance*, 18 (2024), n 1, pp. 3-32, <https://doi.org/10.1111/rego.12512>.

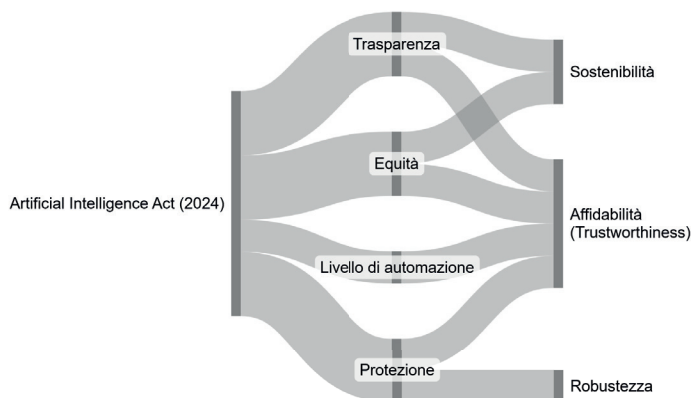


Figura 2. Corrispondenza tra i principi e le proprietà che caratterizzano i sistemi di IA simbiotica conformi con l'AI Act

4. Indicazioni pratiche

L'obiettivo di creare sistemi di IA che riescano a stabilire una relazione *simbiotica* con gli esseri umani può risultare una grande sfida per progettisti e sviluppatori, i quali devono porre l'attenzione su molteplici aspetti restando fedeli all'approccio centrato sulle persone (dall'inglese, *human-centered*). L'IA simbiotica, infatti, necessita di un approccio multidisciplinare per essere costruita, coinvolgendo l'interazione persona-macchina, l'ingegneria del *software*, l'intelligenza artificiale, la legge e l'etica²³. Tale approccio ha guidato la costruzione del *framework FrameSAI* (dall'inglese: *Framework to create Symbiotic Artificial Intelligence*)²⁴ che ha l'obiettivo di stabilire un approccio sistematico e con diversi livelli di granularità per creare sistemi di IA simbiotici. Questo *framework*, illustrato nella fig. 3, pre-

²³ G. Desolda *et al.*, "From Human-Centered to Symbiotic Artificial Intelligence", cit.

²⁴ M. Calvano *et al.*, "Guiding Designers and Developers in Creating Symbiotic AI Systems", in *Proceedings of the 16th Biannual Conference of the Italian SIGCHI Chapter*, ACM, 6 ottobre 2025, pp. 1-5, <https://doi.org/10.1145/3750069.3750095>.

senta una struttura a tre livelli: *Principi*, *Linee Guida* e *Criteri di Successo*. Come suggeriscono questi nomi, i tre livelli sono stati realizzati con l’obiettivo di raggiungere diversi livelli di astrazione, partendo da un’alta generalità dei principi per finire alla pragmaticità e concretezza dei criteri di successo.

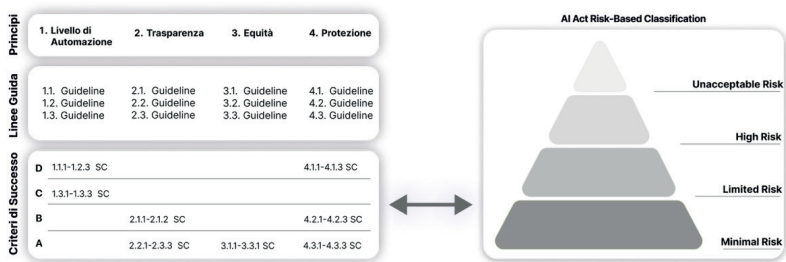


Figura 3. Struttura di FrameSAI con i tre livelli e mappatura con l’AI Act

Nello specifico, questa sezione presenta le linee guida e i criteri di successo che consentono di creare sistemi di IA simbiotica conformi all’AI Act.

Le linee guida definiscono le caratteristiche che il sistema di IA deve possedere e il comportamento che deve manifestare per favorire una collaborazione efficace con gli esseri umani. In particolare, sono state definite 12 linee guida, ciascuna associata a uno o più principi o linee guida e presenta la seguente struttura: <Soggetto> deve <obbligo> <azione> [<motivazione>] (<Principio/i>), dove:

- *Soggetto* può essere un modello di IA, un sistema di IA o un umano;
- *Obbligo* consiste nel requisito o la caratteristica da soddisfare;
- *Azione* riguarda il comportamento che il soggetto deve manifestare;
- *Motivazione* è un parametro opzionale che arricchisce la linea guida, fornendo ulteriori dettagli sui motivi per i quali l’azione o l’obbligo devono essere soddisfatti da parte del soggetto;
- *Principio/i* riporta il singolo o più principi da cui la linea guida deriva.

I criteri di successo sono istruzioni verificabili che derivano da una specifica linea guida e consentono di comprendere se il sistema è conforme a essa. Nello specifico, i criteri di successo sono enunciati con un basso livello di generalizzabilità rispetto alle linee guida, dettagliando le istruzioni in base alla tecnologia, al contesto d'uso e al livello di rischio del sistema. Essi sono stati definiti con la seguente struttura: *<Soggetto> deve <obbligo> <azione>*, dove:

- *Soggetto* può essere un modello di IA, un sistema di IA, un umano o l'*output* generato dall'IA;
- *Obbligo* consiste nel requisito o la caratteristica da soddisfare;
- *Azione* riguarda il comportamento che il soggetto deve manifestare;

La struttura dei criteri di successo è stata definita a partire da quella delle linee guida, ma con un livello di specificità maggiore in cui è importante comprendere anche quali sono le caratteristiche dell'*output* generato dall'IA.

4.1. Esempi e applicazione delle linee guida

Alcuni esempi di linee guida sono:

(G1) *“L'essere umano deve poter verificare, monitorare e supervisionare il comportamento del sistema di IA, tenendo conto di come esso funziona e delle decisioni che prende”* (Livello di Automazione, Trasparenza).

(G4) *“L'essere umano deve essere guidato verso comportamenti etici durante la riconfigurazione del modello di IA, al fine di evitare danni a entrambe le parti”* (Equità, Livello di Automazione).

Nel caso di (G1), la linea guida esplicita che il sistema di IA deve dare la possibilità all'essere umano di poter comprendere il processo decisionale del modello attraverso meccanismi che vedono coinvolti sia i paradigmi dell'Interazione Persona-Macchina che i tecnicismi appartenenti all'Intelligenza Artificiale²⁵. Per ri-

²⁵ A. Esposito *et al.*, “The Fine Line between Automation and Augmentation in Website Usability Evaluation”, in *Scientific Reports*, 14 (2024), n. 10129, <https://doi.org/10.1038/s41598-024-59616-0>; R.A. Calvo *et al.*, “Supporting Human Autonomy in AI Systems: A Framework for Ethical Enquiry”,

trovare l'applicazione concreta di questa indicazione, si riporta un esempio nell'ambito medico: dominio in cui questi sistemi sono spesso progettati con l'obiettivo di classificare dati per supportare i medici nel processo diagnostico. Si fornisce il seguente scenario: un modello viene utilizzato per classificare delle risonanze magnetiche del cervello per determinare se il paziente è affetto da un tumore generando una spiegazione esaustiva e comprensibile che l'utente possa usare come strumento per valutare la qualità della risposta dell'IA ed effettuare la decisione. Nello specifico, la spiegazione indica quali caratteristiche del cervello hanno influenzato l'*output* dell'IA per poi generare la diagnosi. Tale aspetto è molto importante poiché se la spiegazione è presentata in un formato non familiare al medico o che non può comprendere, nonostante sia matematicamente e concettualmente corretta, il risultato finale potrebbe essere fortemente influenzato negativamente. Al contrario, se il sistema è stato creato coinvolgendo l'utente finale e, dunque, conoscendo le modalità più adatte e consone per la presentazione della spiegazione, il processo di interazione risulterà più efficace²⁶. In altre parole, costruire un sistema di IA senza conoscere le caratteristiche dell'utente finale, potrebbe portare ad effettuare decisioni sbagliate con potenziali gravi conseguenze, specialmente in domini ad alto rischio.

Mentre (G1) sottolinea caratteristiche più architetturali, (G4) pone l'attenzione sulla relazione inevitabile tra i meccanismi d'interazione dei sistemi e l'eticità degli stessi. Infatti, con questa linea guida si indica che il sistema deve guidare l'utente nel processo di riconfigurazione (ovverosia richieste di modifiche al funzionamento del sistema) affinché questi possa raggiungere i suoi obiettivi in maniera corretta e senza causare danni a se stes-

in C. Burr, L. Floridi (eds.), *Ethics of Digital Well-Being*, vol. 140, Philosophical Studies Series, Springer International Publishing, Cham, 2020, https://doi.org/10.1007/978-3-030-50585-1_2; B. Shneiderman, *Human-Centered AI*, cit.; U. Farooq *et al.*, "Human Computer Integration versus Powerful Tools", in *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, ACM, 6 maggio 2017, pp. 1277-1282, <https://doi.org/10.1145/3027063.3051137>.

²⁶ G. Desolda *et al.*, "A Human-AI Interaction Paradigm and Its Application to Rhinocytology", in *Artificial Intelligence in Medicine*, 155 (2024), n. 102933, <https://doi.org/10.1016/j.artmed.2024.102933>.

so o al sistema che sta utilizzando²⁷. Per coerenza con l'esempio precedente, si ripropone lo stesso scenario dell'IA applicata alla medicina per illustrare un esempio di corretta applicazione della linea guida. Se un medico non è d'accordo con l'*output* del sistema, nonostante abbia analizzato la spiegazione e correttamente interpretato il processo di ragionamento del modello, deve avere la possibilità di cambiare il risultato fornendo il suo riscontro. In questo caso, il medico dovrà essere guidato dal sistema per poter segnalare l'errore riscontrato al sistema e, eventualmente, accompagnare questo *feedback* con una motivazione. Se il riscontro del medico viene utilizzato per migliorare il modello, questi deve esserne consapevole prima di completarne il processo, in modo tale che sappia che le sue azioni avranno conseguenze sul sistema che sta usando e su eventuali diagnosi future²⁸.

4.2. Esempi e applicazione dei criteri di successo

In maniera analoga alle linee guida, due esempi di criteri di successo per (G1) sono:

(SC1) Il sistema mostra la spiegazione corrispondente al processo decisionale del modello di IA.

(SC2) Il sistema mostra una spiegazione che può essere sia testuale che visiva, in contesti ad alto rischio in cui l'utente deve avere una visione complessiva e dettagliata del processo decisionale del modello di IA.

Considerando i due criteri di successo, è possibile constatare un livello di granularità maggiore rispetto alle linee guida rappresentato dall'enfasi posta sugli elementi specifici da conside-

²⁷ G. Pavlidis, "Unlocking the Black Box: Analysing the EU Artificial Intelligence Act's Framework for Explainability in AI", in *Law, Innovation and Technology* 16 (2024), n. 1, pp. 293-308, <https://doi.org/10.1080/17579961.2024.2313795>; A. Năstasă *et al.*, "Artificial Intelligence: Friend or Foe? Experts' Concerns on European AI Act", in *Economic Computation and Economic Cybernetics Studies and Research*, 57 (2023), n. 3, pp. 5-22, <https://doi.org/10.24818/18423264/57.3.23.01>.

²⁸ A. Curci, A. Esposito, "An AI Act-Driven Design for Detecting Brain Tumors through Reconfiguration", in D. Dell'Anna *et al.* (eds.), *Proceedings of the Workshops at the Fourth International Conference on Hybrid Human-Artificial Intelligence*, vol. 4074, CEUR Workshop Proceedings, 2025, <https://ceur-ws.org/Vol-4074/short9-6.pdf>.

rare (ad esempio, la spiegazione) e dalle corrispondenti caratteristiche (ad esempio, visuale o testuale).

Considerando il primo criterio di successo (SC1), si fa riferimento alla possibilità di visualizzare la spiegazione del processo decisionale consentendo agli utenti di comprendere come e perché il modello abbia prodotto un determinato risultato. Inoltre, il secondo criterio di successo (SC2) fornisce un livello di dettaglio maggiore riguardo alle caratteristiche della spiegazione affinché gli utenti possano comprendere a pieno la motivazione per cui è stato fornito loro un determinato *output*.

Ad esempio, nel dominio medico e nello stesso scenario descritto per la linea guida (G1), tali criteri di successo possono considerarsi soddisfatti se il sistema IA fornisce al medico una spiegazione comprensibile e rappresentativa del processo decisionale. In particolare, la spiegazione deve indicare quali caratteristiche del cervello hanno portato alla diagnosi, la quale deve essere supportata da una rappresentazione visiva, come ad esempio una risonanza magnetica del cervello, che evidenzia le aree rilevanti considerate dal modello di IA durante il processo decisionale.

4.3. Processo iterativo di definizione

Come menzionato nei paragrafi precedenti, le linee guida sono state definite in modo iterativo ed empirico. Una prima fase ha riguardato la decomposizione dei principi in istruzioni atomiche con il fine di costruire le radici delle linee guida. A partire da questa versione preliminare, sono stati analizzati i risultati di alcuni esperimenti eseguiti sull'applicazione dell'IA simbiotica in scenari reali. Nello specifico, è stato considerato un sistema applicato nel campo medico per supportare rinocitologi nella classificazione di cellule attraverso l'utilizzo dell'IA. I risultati ottenuti sono stati utilizzati per arricchire le linee guida con il fine di renderle più complete ed esplicative, arrivando ad una prima versione stabile. Successivamente, è stato necessario valutarle in termini di comprensibilità e rilevanza rispetto ai pilastri dell'IA simbiotica. Pertanto, è stato eseguito uno studio pilota con due ricercatori in Interazione Uomo-Macchina ed Intelligenza Artificiale, chiedendo loro di valutare ogni linea

guida assegnando un punteggio nell'intervallo 1-5, dove 1 rappresenta il giudizio più negativo e 5 quello più positivo. Inoltre, per avere una visione più ampia riguardo ai punteggi inseriti, è stato chiesto ai ricercatori di motivare il punteggio assegnato e di esplicitare il cambiamento richiesto. Chiedendo ai ricercatori di commentare e spiegare la loro valutazione con dei commenti, è stato possibile migliorare le linee guida, generando una nuova versione. Questo studio ha generato risultati promettenti, con riscontri positivi rispetto alla comprensibilità e alla struttura delle linee guida. Di seguito, si riporta un esempio di cambiamento per (G5) che originariamente indicava *“L'essere umano non deve essere manipolato dal sistema di IA attraverso comportamenti persuasivi o ingannevoli”*; i ricercatori hanno riscontrato ambiguità e poca chiarezza rispetto ai casi in cui l'umano abbia intenzioni negative e non etiche nel momento in cui interagisce con un sistema di IA. Pertanto, è stata modificata come segue: *“L'essere umano non deve essere manipolato dal sistema di IA attraverso comportamenti persuasivi, a meno che non debba essere dissuaso da intenzioni non etiche”*²⁹.

Quest'ultima è in fase ulteriore di valutazione, rispetto agli stessi criteri, attraverso uno studio più estensivo e approfondito; al momento della scrittura di questo capitolo, i dati raccolti sono in fase di analisi per la delineazione di un'ultima e definitiva versione delle linee guida con il fine di dare un risvolto più pratico a questa ricerca. Nello specifico, si definiranno indicazioni concrete per la progettazione – attraverso dei *Design Pattern* – e per la valutazione – con i *Criteri di Successo* – sottolineando l'approccio *top-down* e mappando le “foglie” (nello specifico, i criteri di successo) del *framework FrameSAI* alla sua “radice”: una relazione simbiotica tra umani e tecnologia.

5. Conclusioni

L'evoluzione dell'IA e la sua pervasività in numerosi ambiti della società richiedono la creazione di sistemi conformi alla legge e in grado di tutelare i diritti fondamentali degli esseri

²⁹ M. Calvano *et al.*, “Guiding Designers and Developers in Creating Symbiotic AI Systems”, cit.

umani e l'ambiente che li circonda. In questo contesto, l'AI Act rappresenta un passo fondamentale per garantire che lo sviluppo e l'utilizzo dei sistemi di IA avvengano in modo etico e responsabile senza recare danno. Per raggiungere questi obiettivi, è necessario impiegare un approccio *centrato sulle persone* per creare sistemi di IA focalizzati sulle caratteristiche e necessità degli utenti finali, considerandoli non solo come utenti passivi ma come partecipanti attivi ai processi decisionali dell'IA.

In questo contesto, l'IA simbiotica, una specializzazione dell'IA *centrata sulle persone*, promuove un rapporto di collaborazione e continuo miglioramento reciproco tra le due parti. *FrameSAI* costituisce un contributo significativo per la progettazione di sistemi di IA simbiotica conformi all'AI Act, fornendo uno strumento capace di integrare principi, linee guida e criteri di successo che, attraverso diversi livelli di astrazione, possono guidare i progettisti nella creazione di sistemi intelligenti simbiotici.