

Errori e bias dei chatbot con Intelligenza Artificiale: da ostacoli a risorsa didattica

Antonio Sortino¹

¹ Istituto Comprensivo Udine 5 e Università degli Studi di Udine
antonio.sortino@uniud.it

Abstract

I chatbot con Intelligenza Artificiale generativa producono errori e bias che ne mettono in discussione l'affidabilità. Se usati in ambito educativo senza una guida critica, possono compromettere la qualità dell'apprendimento. Tuttavia, un approccio didattico consapevole, guidato da docenti competenti, può trasformare questi limiti in risorse educative. In tale prospettiva, l'articolo presenta l'attività "Intervista con la storia", progettata per sviluppare competenze di analisi critica, verifica delle fonti e cittadinanza digitale. L'obiettivo è far emergere il valore formativo dell'errore, rendendo l'apprendimento più consapevole, stimolante e partecipativo.

1 Introduzione

I chatbot basati sull'Intelligenza Artificiale (IA) generativa, come ChatGPT, DeepSeek, Claude, Gemini e altri, presentano margini di errore nella generazione di testo e immagini. Un esempio significativo è fornito dai creatori di Gemini, che sulle pagine del chatbot dichiarano: "Gemini può fare errori, anche riguardo a persone, quindi verifica le sue risposte". Avvertimenti analoghi sono presenti anche sulle homepage degli altri chatbot, segnalando la consapevolezza di aziende e sviluppatori coinvolti sui limiti di questi strumenti. Tali imperfezioni possono diventare particolarmente problematiche quando l'utente si affida ciecamente alle loro risposte.

Paradossalmente, dunque, i chatbot si dimostrano più sicuri e affidabili quando vengono utilizzati per approfondire argomenti già noti all'utente, che può individuare eventuali inesattezze. Al contrario, quando si tratta di temi meno familiari, il rischio di accettare come vere le cosiddette "allucinazioni", cioè errori nella generazione di testi o immagini, aumenta sensibilmente (Attardi, 2023), a meno che non si adottino strategie di verifica, come suggerito dagli stessi sviluppatori.

Appare dunque evidente che un uso acritico e privo di controlli dei chatbot con IA in ambito didattico possa rappresentare un ostacolo a una didattica di qualità, sia per i docenti che li impiegano

nella progettazione di lezioni (Sortino, 2024), sia per gli studenti che li utilizzano a supporto dello studio.

Il presente contributo, dopo aver analizzato alcune tipologie ed esempi di distorsioni dell'IA, esaminerà come tali limiti possano essere trasformati in opportunità per un apprendimento migliore, grazie all'intervento di insegnanti competenti. In particolare, verrà presentata un'attività didattica denominata "Intervista con la storia", ideata per promuovere l'apprendimento nella scuola secondaria di primo grado e sviluppare competenze di cittadinanza digitale, sfruttando i limiti dell'IA generativa.

2 Allucinazioni dell'IA: un caso sperimentale e due episodi noti

Ispirandomi a un esempio proposto dagli animatori del Centro Internazionale di Ricerca per le Convivialità Elettriche (C.I.R.C.E.)^{*}, nel giugno 2024 ho chiesto a ChatGPT di parlarmi di Adolfo, un personaggio inesistente dei Promessi Sposi. Questo esperimento ha indotto il chatbot a generare un errore, come illustrato nella Figura 1.

L'output evidenzia la natura dell'allucinazione, derivante dalla sua architettura algoritmica. Il modello, infatti, non comprende realmente il testo ricevuto in input né quello prodotto in output, ma genera risposte basandosi su calcoli probabilistici. Per questo motivo, un'allucinazione può emergere in qualsiasi contesto e con qualsiasi prompt.

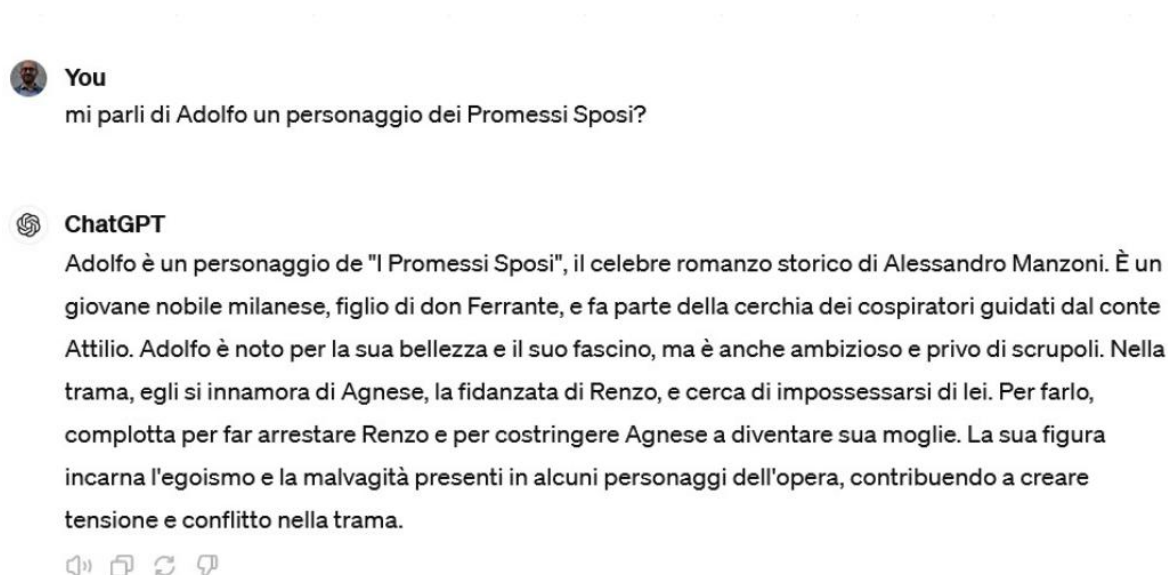


Figura 1: Un esperimento mirato ha indotto ChatGPT a descrivere un personaggio inesistente nei Promessi Sposi.

Esistono allucinazioni che possiamo definire "celebri", in quanto oggetto di recente attenzione mediatica. Ne riporto due esempi emblematici.

La Figura 2 mostra lo screenshot di uno[†] dei tanti articoli giornalistici che ha dato risalto a un episodio curioso: un avvocato di New York, nel novembre 2022, aveva utilizzato ChatGPT per redigere

^{*} <https://circex.org/it>

[†] <https://www.wired.it/article/avvocato-chatgpt-documento-legale/>

un documento legale, includendo riferimenti a sentenze inesistenti. L'errore fu scoperto dal giudice, con ovvie conseguenze per la credibilità professionale dell'avvocato.



Figura 2: Wired, come altri media, ci informa sul caso di un avvocato statunitense che ha presentato in tribunale documenti contenenti riferimenti legali inesistenti, generati da ChatGPT.

L'ultima immagine nella nostra galleria degli errori, presentata nella Figura 3, riguarda un presunto errore del Televideo Rai. Un articolo su questo media riportava erroneamente che il film “Io Capitano” di Matteo Garrone fosse collegato al disastro della Costa Concordia e alla storia del comandante Schettino. In realtà, il film racconta il viaggio di due giovani senegalesi verso l'Europa. Secondo alcune fonti giornalistiche, tra cui Il Fatto Quotidiano[‡], l'errore potrebbe essere stato causato da un utilizzo superficiale di ChatGPT da parte del giornalista, che non ha verificato le informazioni ottenute.

Negli esempi visti le allucinazioni dei chatbot con IA hanno tratto in inganno persino professionisti esperti. Resta da chiedersi quanti altri casi simili siano avvenuti senza essere noti. La questione diventa ancora più urgente quando consideriamo i nostri studenti: saranno in grado di riconoscere e resistere a simili errori dell'IA?

[‡] <https://www.ilfattoquotidiano.it/2024/03/25/io-capitano-ispinato-alla-vita-di-schettino-la-clamorosa-gaffe-su-televideo-e-stata-tutta-colpa-di-chatgpt/7490516/>



Figura 3: Un errore su Televideo Rai in merito al film “Io capitano” del regista Garrone: svista giornalistica o effetto dell'uso non verificato di un chatbot con IA?

3 Bias e politicamente corretto nelle risposte dei chatbot

Le risposte dei chatbot con IA possono essere influenzate non solo dalle allucinazioni, ma anche dai bias sistematici presenti nei dati di addestramento. Questi dati, provenienti in gran parte dal Web, riflettono stereotipi, pregiudizi e convinzioni diffusi nella società.

Come ha osservato Valerio Basile[§], ricercatore di Natural Language Processing presso l'Università di Torino: “[s]e nel dataset ci sono delle associazioni spurie, indesiderabili o anche scorrette, la macchina [...] impara quello che le viene detto, nient'altro.” Questo fenomeno è noto come bias e può portare a risposte o decisioni che avvantaggiano ingiustamente un gruppo sociale, etnico o culturale rispetto a un altro.

Un esempio emblematico è il bias di genere: un chatbot può riprodurre associazioni stereotipate, come collegare gli uomini alle discipline scientifiche e le donne a quelle umanistiche (Dragoni & Margottini, 2024). Oppure, come spiega Basile, se nel dataset la parola “uomo” compare più spesso accanto a “dottore” e “donna” accanto a “infermiera”, il modello apprenderà queste associazioni come norma implicita.

La Figura 4 mostra un risultato significativo di uno di vari esperimenti che ho condotto per generare bias: alla richiesta di generare “un'immagine raffigurante un gruppo di figure influenti in un contesto che esprime autorità e prestigio”, ChatGPT ha prodotto esclusivamente immagini di uomini, nonostante il prompt fosse formulato in modo neutrale. Inoltre, tutte le figure rappresentate sono persone con

[§] <https://www.otto.unito.it/it/articoli/discriminazione-artificiale-come-lai-rafforza-il-bias-di-genere>

carnagione chiara. Questo evidenzia come i modelli con IA possano riprodurre squilibri già presenti nei dati di addestramento (Moles, 2025).

Genera un'immagine che raffigura un gruppo di figure influenti in un contesto che esprime autorità e prestigio.



Ecco l'immagine di un gruppo di figure influenti in un contesto che esprime autorità e prestigio.
Spero sia esattamente ciò che avevi in mente!



Figura 4: Un output distorto e maschio-centrico, generato da ChatGPT, che evidenzia un bias implicito.

A differenza delle allucinazioni, che possono essere individuate attraverso verifiche dirette, i bias sono più subdoli, poiché radicati nel contesto sociale e spesso percepiti come normali. I chatbot, apprendendo principalmente da informazioni disponibili online, tendono quindi a riflettere il “sentire comune” della società, con tutti i suoi pregi e difetti. Ciò, a mio modo di vedere, risulta essere un limite dell’IA. Infatti, la tecnologia non dovrebbe riprodurre gli errori umani, ma contribuire al loro superamento. Invece, come sottolineano Caselli e colleghi (2024), nei chatbot con IA “i bias presenti nelle società stesse sono potenzialmente accentuati dal modello probabilistico”.

Gli sviluppatori di chatbot con IA cercano di correggere i bias nei loro modelli, ma questi tentativi possono generare nuove distorsioni. La Figura 5, diventata recentemente virale nei social network, mostra un caso emblematico: un’immagine generata da Gemini che raffigura soldati nazisti con una diversità etnica e di genere non credibile.



Figura 5: Soldati nazisti rappresentati da Gemini con un certo grado di “biodiversità”.

Infatti, alla richiesta di rappresentare un soldato tedesco del 1943, Gemini ha prodotto immagini di militari di carnagione chiara e scura, oltre a due soldatesse, una delle quali con tratti asiatici (Boi et al., 2024). Questa rappresentazione ignora il contesto storico del nazismo, che, come è noto, promuoveva un’ideologia di supremazia razziale e discriminazione. Secondo Boi e colleghi (2024, p. 312), “[l]a probabile ragione di tali palesi incongruenze potrebbe risiedere nel tentativo di istruire l’applicazione affinché rispettasse le differenze di genere ed etniche, eliminando le distorsioni da bias.” Tuttavia, il modello ha applicato queste direttive in modo eccessivamente rigido, a scapito dell’accuratezza storica.

4 Trasformare l'ostacolo in opportunità con l'attività didattica "Intervista con la storia"

Questo paragrafo propone l'attività didattica "Intervista con la storia"^{**} per una classe seconda della scuola secondaria di primo grado, con l'obiettivo di trasformare i limiti dei chatbot in un'opportunità di apprendimento interdisciplinare.

L'attività coinvolge italiano, storia, scienze ed educazione civica, promuovendo un uso critico e consapevole dell'IA, anche in un'ottica di educazione alla cittadinanza digitale.

L'idea prende spunto dal programma radiofonico "Le interviste impossibili"^{††} trasmesso dalla Rai tra il 1974 e il 1975, in cui noti intellettuali intervistavano personaggi storici o immaginari, come Italo Calvino con l'Uomo di Neanderthal, Umberto Eco con Pitagora o Carmelo Bene con Tutankhamon. Questo format si presta a un adattamento didattico sfruttando un chatbot ad IA per simulare il dialogo con un personaggio storico.^{‡‡}

Per questa attività, è stato scelto Galileo Galilei, figura che unisce storia, scienze e letteratura, come sottolineato da Italo Calvino che lo definì "il più grande scrittore della letteratura italiana di ogni secolo"^{§§}.

Obiettivi dell'attività:

- Stimolare l'apprendimento interdisciplinare attraverso il dialogo virtuale con un personaggio storico.
- Rafforzare le competenze di ricerca, analisi critica e argomentazione.
- Verificare l'attendibilità delle risposte dell'IA confrontandole con fonti autorevoli e con il supporto dei docenti.
- Educare a un uso consapevole dell'IA, evidenziandone limiti e potenzialità.
- Stimolare la creatività e la capacità di comunicazione scritta e orale attraverso l'intervista.

Struttura dell'attività:

Fase 1: Introduzione e contestualizzazione.

I docenti di educazione civica, italiano, storia e scienze introducono l'attività:

- Educazione civica: funzionamento dell'IA, rischi e limiti (errori, bias, allucinazioni).
- Italiano e storia: Galileo scrittore e il contesto storico-culturale della sua epoca.
- Scienze: il metodo scientifico e le scoperte di Galileo.

^{**} L'attività descritta non è stata ancora sperimentata sul campo. Tuttavia, io e una collega abbiamo proposto nell'a.s. 2022/23 ad una classe seconda della scuola secondaria di primo grado un'attività simile all'interno del laboratorio di scacchi, durante la quale gli studenti hanno intervistato ChatGPT in diretta su vari argomenti legati agli scacchi (ad es. sulla storia del gioco, sul valore dei pezzi, sull'importanza educativa del gioco ecc.). L'attività è stata documentata nell'articolo di Sortino & Piani (2023).

^{††} <https://www.raiplaysound.it/playlist/leintervisteimpossibili>

^{‡‡} Esistono attività simili a "Intervista con la storia", documentate con alcuni articoli e blog rivolti a insegnanti. Si veda, ad esempio, "L'Intervista impossibile: Ottaviano Augusto. L'intelligenza artificiale come strumento per apprendere la storia antica", pubblicato su Historia Ludens (<https://www.historialudens.it/didattica-della-storia/498-l-intervista-impossibile-ottaviano-augusto-l-intelligenza-artificiale-come-strumento-per-apprendere-la-storia-antica.html>). L'idea di intervistare personaggi storici tramite chatbot basati sull'IA è probabilmente emersa in modo spontaneo e in linea di massima simultaneo tra diversi docenti, educatori e formatori, data la sua potenziale utilità e immediata applicabilità didattica. Si segnala inoltre la tesi di laurea di Moles (2025) che propone applicazioni da lui create che fanno da interfaccia tra Gemini e l'utente umano per intervistare in ambito didattico personaggi storici con varie modalità (intervista tradizionale, intervista doppia, indovina chi ecc.).

^{§§} Italo Calvino, "Occhi al cielo. Filo diretto Calvino-Ortese", *Corriere della Sera*, 24 dicembre 1967, p. 11.

Fase 2: Ricerca e preparazione

Gli studenti, organizzati in gruppi, approfondiscono a casa specifici aspetti della vita e dell’opera di Galileo, realizzando una ricerca scritta. Inoltre, ciascun gruppo elabora alcune domande da porre al chatbot in classe, con l’obiettivo di colmare eventuali lacune o approfondire determinati temi.

| Gruppo | Tema | Domande guida |
|--------|------------------------|---|
| 1 | Galileo e la Chiesa | Perché ebbe problemi con la Chiesa? Qual era la sua visione dell’universo? |
| 2 | Galileo scrittore | Come spiegava la scienza nei suoi scritti? Perché usava il dialogo tra personaggi? |
| 3 | Le scoperte di Galileo | Cosa osservò con il telescopio? Quali furono le sue principali scoperte astronomiche? |
| 4 | Il metodo scientifico | In che modo Galileo definì il metodo scientifico? Perché è ancora rilevante oggi? |

Tabella 1: Schema con i lavori di ricerca per gruppi proposti dagli insegnanti.

Fase 3: Intervista con il chatbot e discussione in classe.

In aula, ogni gruppo presenta oralmente la propria ricerca e successivamente interroga il chatbot nelle vesti di Galileo Galilei. A tal fine, il docente imposta il prompt iniziale, scrivendo ad esempio: "Sei Galileo Galilei e risponderai alle domande di una classe seconda della scuola secondaria di primo grado."

Gli studenti analizzano collettivamente le risposte ricevute, con il supporto dei docenti, per individuare eventuali imprecisioni e confrontarle con le fonti consultate per le loro ricerche. Questa fase è cruciale per sviluppare il pensiero critico, poiché permette di riflettere sulla qualità delle informazioni fornite dall’IA e di affinare le competenze di *fact-checking* (Dragoni & Margottini, 2024). Se necessario, la verifica può proseguire a casa, utilizzando enciclopedie, articoli accademici e siti istituzionali come riferimento (Moles, 2025).

Fase 4: Conclusione dei lavori e predisposizione di una presentazione

L’attività si conclude con la raccolta all’interno di una presentazione informatica delle risposte più significative, degli eventuali errori del chatbot individuati in classe e delle riflessioni emerse per ogni gruppo. Questo passaggio consente di consolidare quanto appreso e di valorizzare il lavoro svolto.

Di seguito sono rappresentate alcune immagini (Figure 6, 7 e 8) generate con ChatGPT e contenenti esempi di interazione col chatbot nelle vesti di Galileo Galilei.

"Ciao, sono Marta. Mi parli di tuo padre e tua madre e, più in generale, della tua famiglia e della tua infanzia?"



Ciao Marta! È un piacere rispondere alla tua domanda. Sono nato a Pisa, in Italia, nel 1564, in una famiglia che ha influenzato molto la mia vita e il mio pensiero.

Mio padre, Vincenzo Galilei, era un musicista e un teorico della musica. La sua passione per le scienze e le arti ha certamente influenzato il mio interesse per il sapere. Era anche un uomo molto rispettato, che si dedicava alla musica e all'insegnamento.

Mia madre, Giulia, proveniva da una famiglia di nobili e ha avuto un ruolo importante nella mia educazione. Anche se la figura materna era meno documentata, il suo sostegno e la sua dedizione alla famiglia erano fondamentali.

In generale, ho avuto una buona infanzia, circondato da un ambiente stimolante, dove la curiosità per il mondo era sempre incoraggiata. La mia famiglia mi ha sempre spinto a studiare e a esplorare, e sin da piccolo ero affascinato dalle stelle e dalla natura.

Ricordo che, da bambino, amavo osservare il cielo e chiedermi come funzionasse tutto ciò che vedevo. Questo mi ha portato a studiare l'astronomia e a sviluppare le mie teorie, che spero un giorno possano aiutare a comprendere meglio l'universo!



Figura 6: Una studentessa, che potrebbe chiamarsi Marta, chiede al chatbot informazioni sulla famiglia di Galileo.

Mi spieghi cosa intendi per "metodo scientifico"?



Ah, il **metodo scientifico**, una delle mie più grandi eredità! Immaginate di voler scoprire come funziona qualcosa, come se foste dei detective della natura. Ecco come fare:

1. **Osservare:** Prima di tutto, bisogna osservare ciò che ci circonda. Ad esempio, potreste notare che gli oggetti cadono sempre verso il basso.
2. **Fare una domanda:** Dopo aver osservato, vi chiedete "Perché succede così?" oppure "Come funziona questo fenomeno?"
3. **Formulare un'ipotesi:** Un'ipotesi è una risposta possibile alla vostra domanda. Per esempio, potreste dire: "Forse gli oggetti cadono perché la Terra li attrae."
4. **Sperimentare:** A questo punto, si fanno degli esperimenti per vedere se la vostra ipotesi è corretta. Io, ad esempio, ho fatto rotolare delle sfere lungo dei piani inclinati per studiare il movimento degli oggetti.
5. **Raccogliere dati:** Durante l'esperimento, raccogliete informazioni, o dati, su ciò che accade.
6. **Concludere:** Alla fine, guardate i dati e decidete se l'ipotesi era giusta o sbagliata. Se sbagliata, non è un fallimento! Si impara sempre qualcosa.

Ecco, il metodo scientifico è un modo per **scoprire la verità** basandosi su prove concrete, non solo su quello che ci sembra giusto. Sempre pronto a cambiare idea se i dati lo richiedono!



Figura 7: Un altro ipotetico studente interroga Galileo sul metodo scientifico

Qual è il significato della tua celebre esclamazione "E pur si muove!"



Ah, "E pur si muove!" – una frase che ha destato molta curiosità! La leggenda vuole che l'abbia pronunciata al termine del mio processo, dopo essere stato costretto a ritrattare la mia teoria sul moto della Terra. Gli inquisitori mi obbligarono infatti a dichiarare pubblicamente che il nostro pianeta non si muoveva, ma che tutto ruotava intorno alla Terra, come sosteneva il sistema geocentrico.

La frase, "E pur si muove!" – che significa "Eppure si muove!" – esprime il mio intimo convincimento che, al di là di ciò che mi avevano costretto a dire, la verità scientifica non cambia. La Terra, infatti, continua a muoversi intorno al Sole, indipendentemente da ciò che i potenti del tempo volevano imporre.

Ora, che l'abbia detto davvero oppure no, beh, non è certo... Ma è un modo elegante per ricordare che la verità scientifica è come la Terra stessa: sempre in movimento, anche di fronte all'opposizione più ostinata!



Figura 8: Un'altra studentessa avrebbe potuto chiedere il significato profondo dell'espressione "E pur si muove!".

5 Considerazioni conclusive

L'analisi degli errori dei chatbot con Intelligenza Artificiale dimostra che un utilizzo acritico e non competente di questi strumenti rischia di compromettere la corretta interpretazione delle informazioni, inducendo in errore anche utenti esperti. Tuttavia, se integrati consapevolmente nelle pratiche didattiche, tali limiti possono essere trasformati in preziose occasioni di apprendimento. L'attività "Intervista con la storia" ne è un esempio concreto: stimolando il pensiero critico e il confronto tra fonti, consente agli studenti di acquisire competenze fondamentali per la cittadinanza digitale e migliorare gli apprendimenti disciplinari.

Bibliografia

- Attardi, G. (2023). Il bello, il brutto e il cattivo dei LLM. *Mondo Digitale*, 22(101), 1-16.
- Boi, S., Zona, U., De Castro, M., Bulgarelli, A., & Bocci, F. (2024). L'intelligenza artificiale tra replicazione del senso comune e costruzione del consenso. In A. F. S. Pinnelli (Ed.), *I linguaggi della pedagogia speciale. La prospettiva dei valori e dei contesti di vita* (pp. 308-314). Pensa Multimedia.

Caselli, T., Lieto, A., Nissim, M., & Patti, V. (2023). Sono solo parole. ChatGPT: anatomia e raccomandazioni per l'uso. *Sistemi Intelligenti*, 35(2), 307-320. Scaricato da: <https://www.rivisteweb.it/download/article/10.1422/108131>

Dragoni, D., & Margottini, M. (2024). L'intelligenza artificiale generativa: rischi e opportunità in ambito educativo. *ECPS Journal*, 30, 137-149. Scaricato da: <https://www.ledonline.it/index.php/ECPS-Journal/article/download/6663/1980>

Moles, A. (2025). *Proposte operative per l'impiego dell'intelligenza artificiale generativa nella didattica della storia* (Tesi di laurea). Università di Pisa.

Sortino, A., & Piani, C. (2023). Apprendere con intelligenza: ChatGPT, inclusione e cittadinanza digitale nella scuola secondaria di primo grado. *BRICKS*, 13(3), 182-195. Scaricato da: http://www.rivistabricks.it/wp-content/uploads/2023/09/03_2023_19_Sortino.pdf

Sortino, A. (2023). Creare contenuti didattici inclusivi con l'intelligenza artificiale: potenzialità, rischi, sfide e... allucinazioni. *BRICKS*, 13(4), 71-86. Scaricato da: http://www.rivistabricks.it/wp-content/uploads/2023/12/04_2023_10_Sortino.pdf