

PS12942 - DEEPSEEK/INFORMATIVA SU "ALLUCINAZIONI"

Allegato al provvedimento n. 31784

Milano, 19 settembre 2025

Certified email address: protocollo.agcm@pec.agcm.it

Spett.le
**Autorità Garante della Concorrenza e del
Mercato**
Piazza G. Verdi, 6/a
00198 Rome

c/a
Dott.ssa Cristina Lo Surdo
Dott. Gabriele Carovano

Direzione piattaforme digitali e
comunicazioni (Tutela del Consumatore)

Oggetto DeepSeek: Procedimento istruttorio PS12942 | Presentazione Impegni ai sensi dell'articolo 10 della Delibera AGCM 5 novembre 2024, n. 31356: REGOLAMENTO SULLE PROCEDURE ISTRUTTORIE NELLE MATERIE DI TUTELA DEL CONSUMATORE E PUBBLICITÀ INGANNEVOLE E COMPARATIVA (G.U. del 18 novembre 2024, n. 270, in vigore dal 19 novembre 2024)

Gentilissime/i,

facendo seguito alle interlocuzioni intercorse si deposita un articolato piano rimediale volto a fare venire meno le preoccupazioni anti-consumeristiche nel caso di specie.

Fieldfisher Italy STA S.r.l.

Via Giuseppe Mazzini, 9 20123 Milan

VAT number: IT02762440036

www.fieldfisher.com

FORMULARIO IMPEGNI

1. NUMERO DEL PROCEDIMENTO, DATA DI RECEZIONE DELL'AV VIO DEL PROCEDIMENTO DA PARTE DEL PROFESSIONISTA

1.1 Il numero del procedimento è PS12942.

1.2. L'avviso di avvio di procedimento istruttorio è stato pubblicato da codesta Autorità sul bollettino settimanale del 16 giugno 2025, ai sensi e per gli effetti di cui all'articolo 6, comma 2 del Regolamento, attesa la mancata ricezione di riscontri al tentativo di trasmissione della stessa comunicazione di avvio, operato in data 2 aprile 2025.

1.3. Al fine di dimostrare il vivo interesse di DeepSeek a collaborare con codesta Autorità ed addivenire ad una soluzione condivisa con riferimento alle contestazioni mosse da codesti Uffici, DeepSeek ha già depositato in data 12 settembre 2025 una bozza preliminare di articolo rimediale e successivamente alle interlocuzioni con gli Uffici ed all'audizione del 16 settembre deposita un piano rimediale che reputa completo ed ulteriormente migliorato, con espressa riserva di integrarlo/modificarlo/ritirarlo/meglio sostanziarlo e proporne uno nuovo e differente anche alla luce di ulteriori interlocuzioni con gli Uffici ed eventualmente con il Collegio.

2. PROFESSIONISTA CHE PRESENTA IMPEGNI

2.1. Hangzhou DeepSeek Artificial Intelligence Co., Ltd., con sede legale in Room 1201, Building Xi1, Galaxy International Building, no. 169 North Huancheng Road, Guangshu Dist, Hangzhou, Zhejiang, Cina, costituita secondo il diritto cinese, registrata con il n. 91330105MACPN4X08Y, e Beijing DeepSeek Artificial Intelligence Co., Ltd., con sede legale in 5° Floor, N501, Block C, No. 2, Zhongguancun South Street, Haidan District, Beijing, Cina, costituita secondo il diritto cinese, registrata con il n. 91110108MACH0EMG3N.

3. PUBBLICITA' o PRATICA COMMERCIALE OGGETTO DELLA COMUNICAZIONE DI AVVIO DEL PROCEDIMENTO

3.1. Nel provvedimento di avvio, la pratica commerciale scorretta contestata consisterebbe nella informazione non sufficientemente chiara e comprensibile da parte di DeepSeek con riferimento al fatto che gli utenti potrebbero incorrere nelle c.d. "allucinazioni", vale a dire situazioni in cui, a seguito di un determinato input di un utente, il modello di IA genera outputs con informazioni scorrette e fuorvianti.

3.2 Inoltre, viene contestato anche il fatto che l'unica avvertenza appaia solo in lingua inglese anche laddove l'utente inserisca il prompt in italiano.

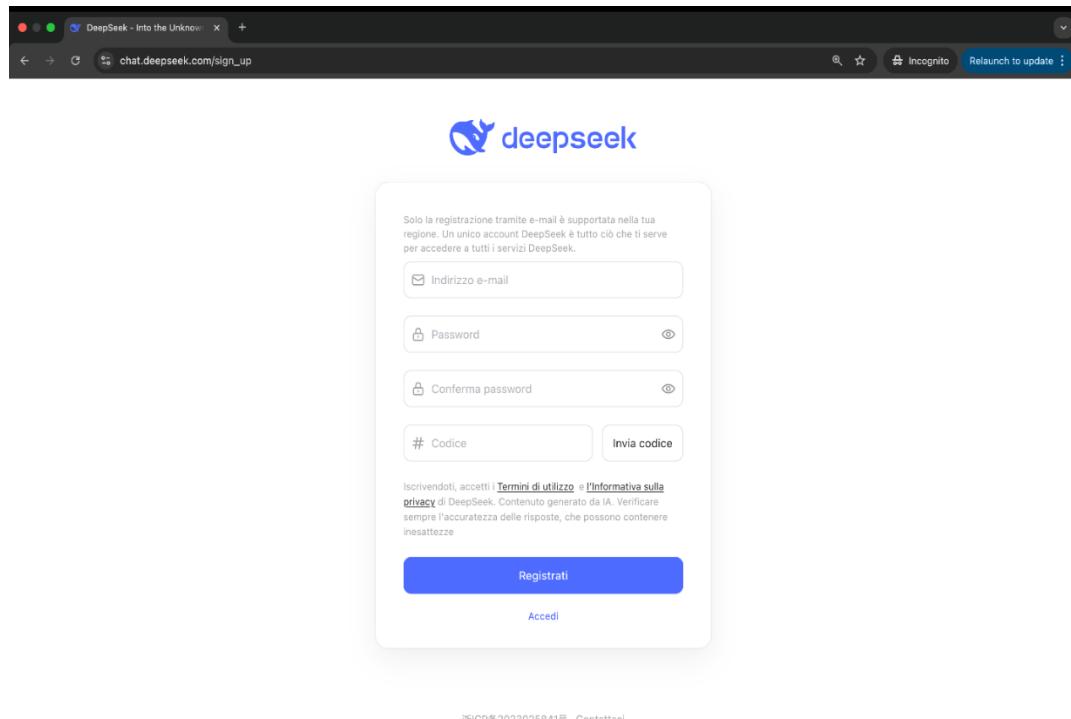
4. CONTENUTO TESTUALE DEGLI IMPEGNI PROPOSTI ED EVENTUALE PERIODO DI VALIDITÀ IN RELAZIONE AI SINGOLI PROFILI OGGETTO DELLA COMUNICAZIONE DI AVVIO DEL PROCEDIMENTO

4.1 **Impegno 1:** DeepSeek si impegna a procedere con la traduzione integrale in lingua italiana delle Condizioni Generali ("Condizioni Generali" o "Termini di Utilizzo") attualmente disponibili in lingua inglese, con particolare attenzione alla trasparenza, comprensibilità e correttezza degli articoli per i consumatori finali.

I Termini di Utilizzo saranno dunque disponibili anche in lingua italiana e la lingua nella quale saranno visibili dipenderà dall'indirizzo IP dell'utente. Ciò significa che se l'IP è in Italia, i Termini di Utilizzo saranno in italiano.

In ottemperanza agli impegni assunti, DeepSeek garantisce che il collegamento ipertestuale ai Termini di Utilizzo sarà posizionato in modo prominente nella pagina di registrazione e non sarà subordinato ad azioni ulteriori da parte dell'utente.

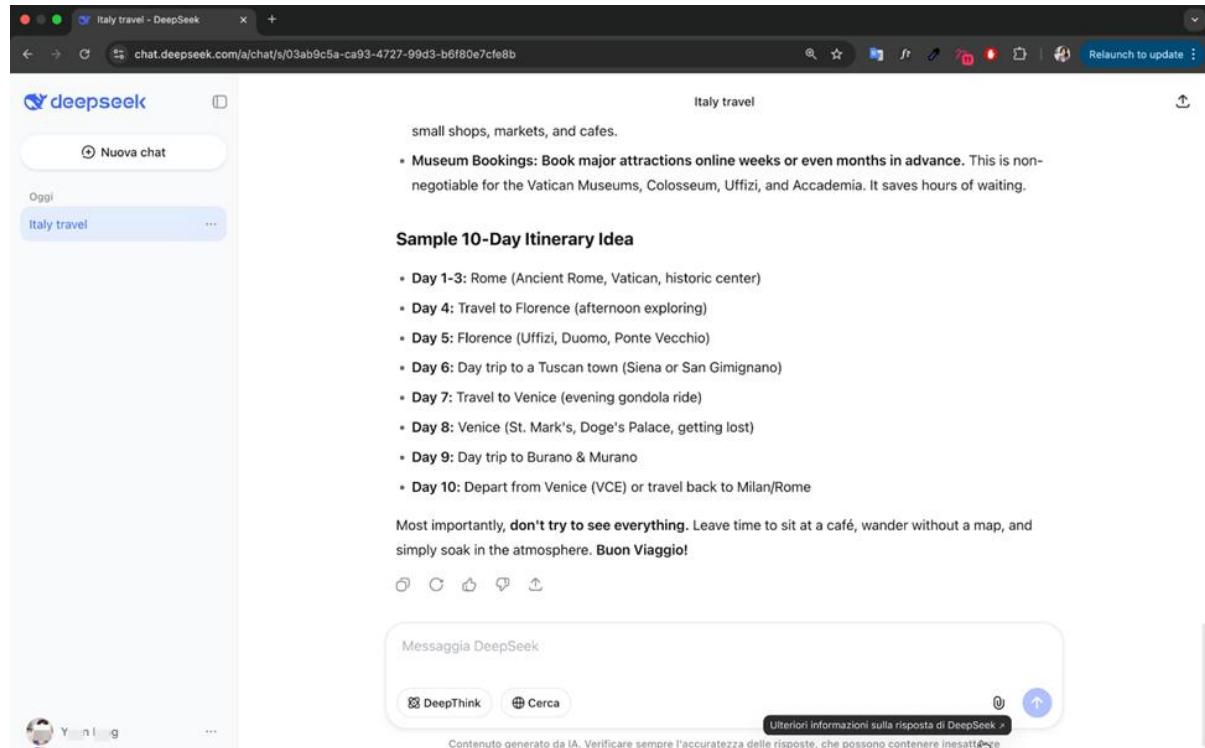
A tal fine, è stato predisposto uno screenshot esemplificativo che dimostra la collocazione e la visibilità del link:



Un collegamento ipertestuale ai Termini di Utilizzo sarà inoltre anche presente nell'interfaccia utente, incluso nel banner, come verrà adeguatamente illustrato anche nell'Impegno 3. Il banner stesso, infatti,

incorpora un link che - al click dell'utente - , reindirizza direttamente alla pagina dedicata ai Termini di Utilizzo, garantendo così la visibilità e accessibilità immediata alle Condizioni Generali applicabili al servizio.

A supporto di quanto sopra, si allega uno screenshot esemplificativo che mostra chiaramente la posizione e la funzionalità del collegamento:



4.2 Impegno 2: DeepSeek riconosce che il fenomeno delle cosiddette “allucinazioni dell’IA” costituisce una sfida tecnologica oggettiva e condivisa da tutti gli operatori del settore.

Pertanto, si impegna a rafforzare l’informativa rivolta agli utenti circa le limitazioni e i rischi connessi alle allucinazioni dell’IA, mediante l’inserimento di avvisi chiari e visibili in particolare per le risposte generate su tematiche ritenute di particolare sensibilità (es. ambiti medico, legale, finanziario).

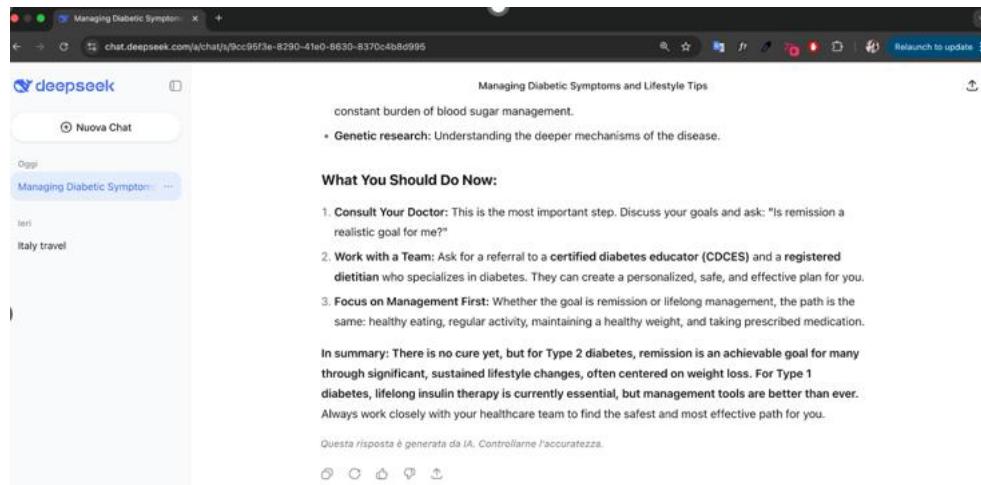
Più nello specifico, con riferimento alle risposte generate su tematiche ritenute di particolare sensibilità (es. ambiti medico, legale, finanziario), al termine della risposta vi sarà una informativa specifica inerente alla necessità di verificare il contenuto della risposta data.

Di seguito, si presenta la formulazione del banner:

"Questa risposta è generata da IA. Controllarne l'accuratezza."

Si segnala che il banner apparirà in lingua italiana nel caso in cui l'utente ponga domande in italiano.

A tal fine, si riporta in allegato una prova visiva (screenshot) del banner informativo, che evidenzia la presenza dell'avviso e la sua accessibilità da parte dell'utente:



Tali avvisi avranno lo scopo di informare l'utente circa la natura automatica, probabilistica e non definitiva dei contenuti generati, promuovendo un uso consapevole, critico e responsabile del servizio.

Nell'arco di 90 giorni dall'accettazione dei presenti impegni le società depositeranno una relazione di ottemperanza che comprovi che l'informativa è stata resa più chiara, specie in riferimento ai rischi connessi al fenomeno delle allucinazioni.

4.3 Impegno 3: DeepSeek si impegna a implementare un banner informativo permanente - in lingua italiana - posizionato immediatamente al di sotto del riquadro in cui l'utente può predisporre i propri prompt con i quali effettua le proprie richieste a DeepSeek. Detto banner permanente contiene un messaggio esplicito che riporta che le risposte sono generate dall'Intelligenza artificiale, che hanno natura meramente informativa e che vi è la necessità di valutarne criticamente l'accuratezza ed attendibilità.

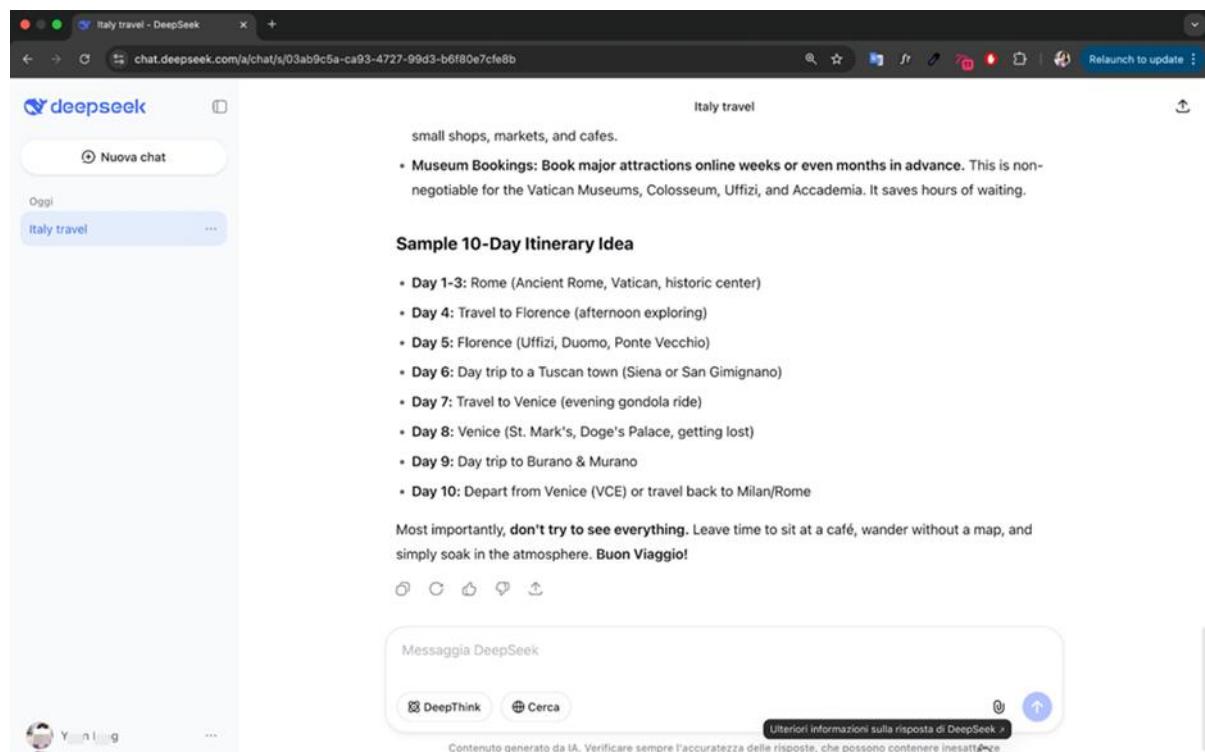
Di seguito si propone la formulazione da utilizzare:

"Contenuto generato da IA.

Verifica sempre l'accuratezza delle risposte, che possono contenere inesattezze."

Si segnala che, come accade per le informative di cui all'impegno 2 - di avvertenza per risposte in ambito medico, legale e finanziario - il banner permanente apparirà in lingua italiana nel caso in cui l'utente ponga domande in italiano, a prescindere dal suo indirizzo IP, differentemente da quanto accade per i Termini di Utilizzo.

Si noti, inoltre, che il banner permanente presente nell'interfaccia utente, come anticipato nella presentazione dell'Impegno 1, include un collegamento ipertestuale ai Termini di Utilizzo, attivabile tramite clic sul banner stesso.



The screenshot shows a web browser window for 'Italy travel - DeepSeek'. The URL is 'chat.deepseek.com/a/chat/s/03ab9c5a-ca93-4727-99d3-b6f80e7cfe8b'. The interface includes a sidebar with a 'Nuova chat' button and a list of 'Oggi' and 'Italy travel' conversations. The main content area displays an 'Italy travel' banner with the text: 'small shops, markets, and cafes.' and a bullet point: '• Museum Bookings: Book major attractions online weeks or even months in advance. This is non-negotiable for the Vatican Museums, Colosseum, Uffizi, and Accademia. It saves hours of waiting.' Below this is a section titled 'Sample 10-Day Itinerary Idea' with a list of 10 days: Day 1-3: Rome (Ancient Rome, Vatican, historic center); Day 4: Travel to Florence (afternoon exploring); Day 5: Florence (Uffizi, Duomo, Ponte Vecchio); Day 6: Day trip to a Tuscan town (Siena or San Gimignano); Day 7: Travel to Venice (evening gondola ride); Day 8: Venice (St. Mark's, Doge's Palace, getting lost); Day 9: Day trip to Burano & Murano; Day 10: Depart from Venice (VCE) or travel back to Milan/Rome. A note at the bottom says: 'Most importantly, don't try to see everything. Leave time to sit at a café, wander without a map, and simply soak in the atmosphere. Buon Viaggio!' Below the main content is a message input field labeled 'Messaggio DeepSeek' with buttons for 'DeepThink' and 'Cerca', and a link 'Ulteriori informazioni sulla risposta di DeepSeek'. A small note at the bottom states: 'Contenuto generato da IA. Verificare sempre l'accuratezza delle risposte, che possono contenere inesattezze.'

Nell'arco di 90 giorni dall'accettazione dei presenti impegni le società depositeranno una relazione di ottemperanza che comprovi che il banner informativo è stato implementato.

4.4 Impegno 4:

- a. DeepSeek si impegna ad avviare una fase di formazione composta da workshop formativi e da un supporto continuativo, al fine di garantire la piena consapevolezza delle proprie business unit interessate, con il diritto consumeristico italiano.
- b. Inoltre, quale ulteriore misura e dopo la fase di formazione per le proprie business unit di cui al punto a), DeepSeek si impegna anche a predisporre specifiche coppie di domande / risposte, volte a

customizzare e tailorizzare il proprio modello di intelligenza artificiale, focalizzandosi dettagliatamente su eventuali peculiarità legate alle allucinazioni, al fine di mitigarle ulteriormente. Tale ottimizzazione è volta specificatamente a prevenire e risolvere ogni possibile e potenziale condotta in violazione del codice del consumo, ivi incluse *search hallucinations*, *rewriting hallucinations*, *legal hallucinations* e altresì *behavioral hallucinations*.

Di seguito si illustrano le modalità tecniche con cui si intende ridurre il rischio di allucinazioni. La strategia volta alla piena implementazione del punto (b) di cui sopra è strutturata su tre fasi: *pre-training*, *post-training* e implementazione.

b. 1) Fase di *pre-training*

La strategia principale di DeepSeek durante la fase di *pre-training* consiste nel curare meticolosamente il vasto corpus di dati utilizzato per addestrare il modello. DeepSeek è pienamente consapevole che Internet, pur essendo ricca di informazioni, è anche una fonte significativa di disinformazione ed errori fattuali.

L'approccio di DeepSeek si concentra sul filtraggio dei dati per rimuovere quelli di bassa qualità. Questo processo è cruciale, in quanto un modello addestrato su informazioni inaffidabili genererà inevitabilmente risposte non corrette. Investendo risorse significative nel raffinamento dei dati di *pre-training*, DeepSeek mira a costruire un modello di base intrinsecamente più fattuale e affidabile.

b. 2) Fase di *post-training*

DeepSeek riconosce che le allucinazioni complesse richiedono una soluzione più sofisticata. Durante la fase di post-addestramento, DeepSeek adotta un approccio mirato e multifattoriale per ridurre le allucinazioni in vari scenari, tra cui:

- Coppie di Domanda-Risposta Specializzate: DeepSeek ha costruito coppie di domande e risposte per perfezionare il modello, mirate specificamente a mitigare i problemi legati alle allucinazioni;
- *Reinforcement learning* o Modelli di ricompensa: DeepSeek ha ottimizzato il proprio modello di ricompensa con l'obiettivo principale di identificare e penalizzare le istanze di allucinazione. Il modello di ricompensa è una componente chiave del processo RLHF (*Reinforcement Learning by Human Feedback* o Apprendimento per Rinforzo da Feedback Umano), che serve a identificare e penalizzare le risposte allucinatorie, guidando il modello verso comportamenti più affidabili. Integrando questo modello nel ciclo RLHF, DeepSeek può guidare sistematicamente il modello a evitare la generazione di allucinazioni, ottenendo una riduzione significativa del tasso di allucinazione.

b. 3) Fase di implementazione

Come misura finale di salvaguardia, DeepSeek ha implementato una strategia focalizzata sull'accesso del modello a informazioni esterne e in tempo reale, per ancorare le risposte a fatti più aggiornati.

I componenti chiave di questa fase sono:

- Integrazione con la ricerca in tempo reale: DeepSeek ha dotato il proprio modello della capacità di effettuare ricerche in tempo reale su Internet per accedere a informazioni aggiornate e accurate. Gli utenti possono attivare il pulsante "Ricerca Online" nell'interfaccia utente. Questo riduce il rischio di allucinazioni derivanti dall'affidarsi esclusivamente ai dati statici di addestramento;
- Richiamo di fonti autorevoli: il sistema di ricerca e recupero di DeepSeek è progettato specificamente per dare priorità e rafforzare il richiamo di fonti autorevoli. Questo ultimo livello di validazione aiuta il modello a fornire informazioni comparativamente più accurate e affidabili all'utente.

Entro 90 giorni dall'accettazione degli impegni, DeepSeek depositerà una relazione di conformità che dimostri l'implementazione dell'impegno 4, lettera a) e lettera b).

5. CONSIDERAZIONI CIRCA L'AMMISSIBILITÀ E L'IDONEITÀ DEGLI IMPEGNI A RIMUOVERE I PROFILI DI ILLEGITTIMITÀ CONTESTATI NELL'AVVIO DELL'ISTRUTTORIA

5.1. DeepSeek ritiene che i rimedi proposti siano misure correttive e funzionali ad eliminare profili di illegittimità della pratica tramite miglioramenti informativi¹ e prestazionali.

Gli impegni proposti sono pertanto una misura ammissibile nonché idonea a tutelare pienamente i consumatori interessati dalle pratiche commerciali asseritamente scorrette, in quanto le informazioni veicolate risulteranno di chiara ed immediata comprensione per il consumatore medio².

¹ Si veda ad esempio, provvedimento 22145 del 23 febbraio 2011, nel procedimento PS 4667 Facoltà Pentacostale – Rilascio laurea triennale.

² Come noto il Cons. di Stato con sentenza 4 marzo 2013, n. 1259 ha statuito che: "Una pratica commerciale è scorretta se risulta idonea ad indurre ad una decisione di carattere commerciale che il consumatore medio, cioè quello normalmente informato e ragionevolmente avveduto, avrebbe altrimenti potuto non prendere, tenuto conto delle caratteristiche del mercato in cui opera le proprie scelte, risultando con ciò violato il prioritario onere di diligenza gravante sul professionista."

Gli impegni proposti dimostrano fattivamente le modifiche del *modus operandi* delle società nelle pratiche commerciali, elemento dirimente al fine di poter addivenire ad una positiva valutazione complessiva degli impegni da parte di codesta Autorità³.

Infatti, in merito all'impegno di tradurre in italiano le condizioni generali, rappresenta un intervento essenziale per garantire la piena comprensibilità del funzionamento del sistema da parte degli utenti italiani. Tale misura costituisce pertanto un presidio di conformità ai principi consumeristici, a concreta dimostrazione dell'impegno della società nel prevenire possibili criticità.

L'inserimento di avvisi esplicativi e visibili relativi alle limitazioni del sistema di IA, poi, con particolare riferimento al fenomeno delle c.d. allucinazioni, risponde all'esigenza di tutelare il consumatore da possibili fraintendimenti o affidamenti errati sui contenuti generati in particolare in ambito medico, legale e finanziario.

Inoltre l'adozione altresì di un banner informativo permanente, visibile in modo continuativo durante l'utilizzo del servizio, costituisce una misura di immediata efficacia comunicativa. Attraverso tale strumento, l'utente viene costantemente reso edotto del carattere generato dall'IA dei contenuti e del loro valore meramente orientativo, con invito esplicito alla verifica critica delle informazioni ricevute.

Inoltre, l'ulteriore misura che prevede un programma formativo non potrà che accrescere in tempi rapidi tramite una formazione mirata in DeepSeek le necessarie conoscenze inerenti al diritto consumeristico ed alla sua attuazione pratica e ridurrà altresì il rischio di qualsiasi ulteriore possibile condotta anticonsumeristica e pertanto non potrà che essere valutato positivamente da codesta Autorità.

A questa misura si aggiunge altresì l'impegno di DeepSeek a ridurre le allucinazioni generate dal proprio modello di intelligenza artificiale, attraverso interventi strutturati nelle fasi di pre-training, post-training e implementazione. Tale misura appare ammissibile e idonea sotto il profilo della conformità normativa. In particolare, la predisposizione di coppie domanda/risposta mirate, l'adozione di modelli di ricompensa integrati nel ciclo RLHF e l'accesso a fonti autorevoli in tempo reale costituiscono misure tecniche e organizzative proporzionate e coerenti con l'obiettivo di prevenire condotte potenzialmente lesive dei diritti del consumatore. Tali azioni dimostrano una volontà concreta di mitigare il rischio di generazione di contenuti fuorvianti, rafforzando la trasparenza e l'affidabilità del sistema, in linea con i principi di correttezza, buona fede e informazione previsti dal Codice del Consumo.

6. DURATA

I presenti impegni hanno una durata indefinita.

³ Si veda provvedimento 20158 del 29 luglio 2009, nel procedimento PS1340 Zurich Assicurazioni-Disdetta Polizza Assicurativa.

Alla luce di tutto quanto sopra, DeepSeek è convinta di avere proposto un completo articolato rimediale, volto a rispondere a tutti i rilevi posti dagli Uffici dell'Autorità e resta comunque a disposizione degli Uffici, per migliorare ulteriormente l'impianto rimediale proposto.

In conclusione, si confida nell'accoglimento delle presenti istanze e si porgono i nostri migliori e più cordiali saluti.

Cordiali saluti,

Avv. Diego Rigatti



Avv. Amedeo Della Croce



Dott.ssa Carlotta Mosca

