Linee guida FSA per la gestione dell'intelligenza artificiale

1. Introduzione

L'intelligenza artificiale («IA») è attualmente sulla bocca di tutti e, secondo i produttori di software, viene impiegata in molte applicazioni. Anche l'avvocatura utilizza in misura crescente i sistemi di IA, in particolare tramite i software di traduzione o nell'ambito dell'analisi di grandi quantità di dati finalizzate a indagini interne svolte per la clientela o nei processi di due diligence. inoltre, sempre più spesso si fa uso anche della cosiddetta IA generativa, ossia un'IA che genera essa stessa contenuti come immagini o testi (ad es. per riassumere, correggere o migliorare un testo.

Queste linee guida sono focalizzate principalmente sull'IA generativa. Le raccomandazioni sono però valide per tutte le applicazioni in cui viene utilizzata l'IA.

Come tutte le nuove tecnologie, anche l'IA apre la strada a nuove opportunità, in particolare in termini d'incremento dell'efficienza, ma cela anche dei rischi. Le presenti direttive sono da intendersi quali linee guida per un impiego responsabile dell'IA nella pratica dell'avvocatura, per permettere agli studi legali di emanare le proprie direttive interne per l'utilizzo dei sistemi di IA.

I sistemi di IA sono in rapida evoluzione. Il contenuto delle presenti direttive può diventare pertanto rapidamente obsoleto. Per tale motivo, si rinuncia a fare riferimento a applicazioni di IA specifiche.

2. Rischi legati alla gestione e utilizzo sicuro

2.1. Segreto professionale degli avvocati, protezione dei dati e altri obblighi di confidenzialità

Nella scelta e nell'utilizzo di qualsiasi software e quindi anche delle applicazioni di IA, si devono preliminarmente chiarire le modalità di gestione dei dati immessi nel sistema, in particolare chiarire chi vi ha accesso e dove viene eseguito il relativo salvataggio (intermedio). Va garantita non solo la salvaguardia del segreto professionale, ma anche l'osservanza della legge sulla protezione dei dati («LPD»), nonché quella di ulteriori, eventuali obblighi di confidenzialità. I membri della Federazione Svizzera degli Avvocati devono inoltre garantire il rispetto del Codice di deontologia.

In linea di massima sono ipotizzabili le seguenti possibilità:

1. Installazione e gestione del software all'interno della rete dello studio legale (cosiddetta soluzione *on-premise*) ed è garantito che nessun dato esca da questa rete interna o venga memorizzato al di fuori dell'infrastruttura dello studio legale.
2. Rispetto delle regole sull'outsourcing, se le applicazioni vengono acquistate tramite un provider ed eventualmente utilizzate tramite la sua infrastruttura. Si rinvia in proposito alle Linee guida FSA per l'outsourcing IT e l'uso di servizi cloud (consultabili all'indirizzo https://digital.sav-fsa.ch/it/digitale-kanzlei-nutzung-von-clouddiensten).
3. Dichiarazione di consenso informato e di rinuncia al segreto professionale e alla LPD da parte del cliente.

Al di fuori di queste possibilità, informazioni riservate, segreti aziendali (strategie, dati finanziari ecc.), informazioni o dati personali di collaboratori, clienti o partner commerciali o di altre persone (in qualsiasi formato, ossia anche inclusi foto, video, ecc.) o contenuti protetti dai diritti di proprietà intellettuale, in particolare da diritti d'autore, non possono essere inseriti in sistemi di IA.

2.2. Verifica (indipendente) dei risultati

I sistemi di IA non sono onniscienti né funzionano sempre perfettamente; i loro risultati possono essere errati, insufficienti o incompleti. È dunque estremamente importante verificare in modo indipendente e critico i risultati, il cosiddetto «*output*», ed eventualmente correggerli risp. integrarli.

In proposito è importante sapere che l'IA non è in grado di verificare i risultati che genera. Per esempio, non si può semplicemente interrogare il sistema di IA per sapere se l'output fornito corrisponde alla verità, poiché i sistemi di IA non sono in grado di fornire in proposito risposte attendibili.

Gli errori o risultati errati di applicazioni di IA possono avere in particolare le seguenti cause:

- Allucinazioni, ovvero l'intelligenza artificiale «inventa» semplicemente l'output.
- Informazioni errate o mancanti perché al sistema di IA manca una base di dati affidabile. Un sistema di IA può, ad esempio, essere addestrato su una base di dati risalente al passato e dunque non essere a conoscenza di eventi successivi a tale data. I sistemi di IA, inoltre, sono tanto più inclini ad allucinazioni quanto meno materiale confluisce nell'addestramento relativo a un determinato argomento.
- Sycophancy: un modello di IA adatta le proprie risposte al punto di vista dell'utente, anche se tale punto di vista è oggettivamente distorsivo.

Va pure considerato che i sistemi di IA possono essere prevenuti. Tali cosiddetti bias possono originare dalla serie di dati utilizzata per l'addestramento del sistema di IA, dalle modalità di addestramento, nonché dalle decisioni di modellizzazione prese dai programmatori. Si tratta un aspetto rilevante soprattutto nel contesto dell'utilizzo di sistemi di IA per l'analisi dei dati, ma occorre prestarvi attenzione anche nell'impiego di sistemi di IA generativi.

2.3. Responsabilità

In qualità di mandatario, l'avvocato risponde del coretto adempimento del proprio mandato. Non gli è segnatamente possibile appellarsi a errori commessi dall'IA.

In caso di impiego di sistemi di IA in accordo con la clientela (ad es. per l'analisi di grandi quantità di dati), si raccomanda di affrontare e trattare preventivamente con il cliente il tema della responsabilità per il risultato e, nei limiti di quanto giuridicamente consentito, di concordare una limitazione della responsabilità.

2.4. Diritto d'autore

La questione relativa all'utilizzo di materiale protetto dal diritto d'autore, in particolare per l'addestramento di Large Language Models (LLM) o di generatori di immagini risp. di video, è attualmente oggetto di un dibattito molto controverso. Altresì controverso è se tali dati di addestramento violino i diritti d'autore dei titolari del diritto in assenza del loro consenso risp. se l'utilizzo per scopi di addestramento rappresenti un utilizzo rilevante dal punto di vista del diritto d'autore. Tali questioni sono tuttavia meno importanti nel caso del mero utilizzo di applicazioni di IA; possono però essere rilevanti se gli studi legali addestrano i propri LLM o integrano LLM esistenti.

In assenza di una chiara regolamentazione del loro diritto di utilizzo, i dati di input per l'addestramento di sistemi di IA o per la generazione di output possono eventualmente violare diritti d'autore di terzi.

In linea di principio gli output generati dall'IA non sono considerati una creazione dell'ingegno ai sensi del diritto d'autore. Se un sistema di IA viene impiegato solo per la raccolta delle idee, per una prima bozza o simili, è comunque possibile che successivamente ne scaturisca una creazione dell'ingegno a carattere individuale e quindi che nonostante l'utilizzo di sistemi di IA, nasca un'opera protetta dal diritto d'autore. Non è neppure del tutto escluso che un determinato output generato da IA violi diritti d'autore di terzi poiché l'output risulta troppo simile a un'opera protetta dal diritto d'autore. Si tratta però di una circostanza molto improbabile.

2.5. Obbligo d'informazione

Nelle loro condizioni generali di utilizzo, alcun fornitori prevedono disposizioni che obbligano gli utenti a rendere noto l'impiego dell'IA. Prima di utilizzare un'applicazione di IA, non da ultimo per tale motivo, si dovrebbero pertanto verificare le condizioni di utilizzo del fornitore in questione.

A prescindere da ciò, un obbligo d'informazione potrebbe eventualmente sussistere anche nei casi in cui il cliente/mandante, esige o si attende l'esecuzione strettamente personale del mandato da parte dell'avvocato.

3. Regolamentazione dei sistemi di IA

In Svizzera, nel novembre del 2023 il Consiglio federale ha incaricato il DATEC di presenare entro la fine del 2024 possibili approcci a una regolamentazione dell'IA. L'analisi deve basarsi sul diritto svizzero vigente e mostrare possibili approcci di regolamentazione per la Svizzera che siano compatibili con la legge sull'intelligenza artificiale dell'UE e con la Convenzione sull'intelligenza artificiale del Consiglio d'Europa (vedi sotto). Le esigenze di regolamentazione sono valutate con particolare attenzione al rispetto dei diritti fondamentali. Vengono presi in considerazione anche gli standard tecnici e le ripercussioni finanziarie e istituzionali dei diversi approcci normativi. Il Consiglio federale prevede di conferire nel 2025 un mandato concreto per lo sviluppo di un progetto di regolamentazione dell'IA.

L'UE sta per adottare una delle prime regolamentazioni complete in materia di IA («legge sull'intelligenza artificiale» o «AI-Act»). Non è ancora stato stabilito quando la legge sull'intelligenza artificiale entrerà in vigore e quando si applicherà.

La legge sull'intelligenza artificiale è improntata a un approccio basato sul rischio e suddivide i sistemi di IA in quattro categorie:

1. Sistemi di IA con un livello di rischio inaccettabile: vale il divieto assoluto d'impiego. Rientrano in questa categoria, in particolare, i sistemi di identificazione biometrica in tempo reale e a distanza («Facial Recognition»), salvo che vengano utilizzati per il perseguimento di reati penali e siano soggetti a severe condizioni (ad es. alla preventiva autorizzazione dell'impiego da parte di un'autorità giudiziaria). Rientrano inoltre in questa categoria il cosiddetto «social scoring», ossia la classificazione dell'affidabilità delle persone fisiche in base al loro comportamento sociale, delle caratteristiche personali o della personalità note o previste, a condizione che ciò comporti una discriminazione o uno svantaggio ingiustificati o sproporzionati di tali persone.
2. I sistemi di IA ad alto rischio sono consentiti solo a severe condizioni (gestione del rischio, governance dei dati, documentazione tecnica, obblighi di registrazione, requisiti di trasparenza, vigilanza, precisione, solidità e cibersicurezza, ecc.). Rientrano in questa categoria ad es. sistemi di IA in relazione a candidature, promozioni o disdette inerenti a rapporti di lavoro, alla verifica del diritto all'assistenza sociale, alla valutazione degli allievi

scolastici, agli esami di ammissione all'università o a previsioni sulla recidiva da parte di criminali, ecc.

3. I sistemi di IA con rischi limitati sono ammessi nel rispetto di limitati requisiti di trasparenza (ad es. informazione dell'utente che questi ha a che fare con un sistema di IA o identificazione riconoscibile di contenuti generati con intelligenza artificiale).
4. I sistemi di IA con rischio nullo / basso, come ad es. in relazione ai videogiochi, sono consentiti e rimangono in larga misura non regolamentati.

La legge sull'intelligenza artificiale prevede la creazione di un comitato europeo per l'intelligenza artificiale e di autorità a livello nazionale, che devono anche poter sanzionare le imprese inadempienti. La legge sull'intelligenza artificiale non contempla la possibilità di invocare individualmente in giudizio l'applicazione del diritti che codifica.

Oltre all'UE, anche il Consiglio d'Europa ha elaborato una Convenzione sull'intelligenza artificiale. Il 17 maggio 2024 il Consiglio dei ministri ha emanato la Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law («Convenzione sull'intelligenza artificiale»), disponibile solo in lingua inglese e francese. La Convenzione sull'intelligenza artificiale mira a garantire l'impiego dell'intelligenza artificiale in modo da rispettare i diritti fondamentali, la democrazia e i principi dello Stato di diritto. Essa è aperta alla firma e alla ratifica degli Stati membri e degli altri Stati che hanno partecipato alla sua elaborazione. La Convenzione sull'intelligenza artificiale entrerà in vigore non appena cinque Stati, di cui almeno tre Stati membri, l'avranno approvata.

Allegato: Glossario

AI o Artificial Intelligence	Vedi IA o intelligenza artificiale
Modello black box	Descrive l'impossibilità, risp. la particolare difficoltà, di verificare l'approccio e i percorsi di soluzione dei sistemi di IA. Per lo più si tratta di aspetti non verificabili per gli utenti e difficilmente verificabili persino per i programmatori di un'IA.
Bias	Prevenzione di un sistema di IA sulla base della serie di dati di addestramento, della modalità di addestramento o delle decisioni di modellizzazione dei programmatori.
Legge sull'intelligenza artificiale (P9_TA(2024)0138)	Legge UE sulla regolamentazione dell'IA.
GPT	Generative Pre-trained Transformer
Allucinazioni	Fatti presunti (falsi), non basati su dati o eventi reali, ma presentati come tali.
Input	Dati che vengono immessi in un sistema di IA.
IA o intelligenza artificiale	Per intelligenza artificiale si intende la capacità di una macchina di svolgere attività che normalmente richiederebbero l'intelligenza umana. Tra questi si annoverano la risoluzione dei problemi, l'apprendimento, il riconoscimento vocale, il processo decisionale e molto altro. Gli algoritmi e i sistemi di IA possono analizzare i dati, riconoscere i modelli e sulla base degli stessi fare previsioni o prendere decisioni.
LLM o Large Language Model	Modello di apprendimento automatico in grado di svolgere compiti nell'ambito del Natural Language Processing
NLP o Natural Language Processing	Elaborazione del linguaggio naturale da parte di un sistema di IA
Output	Risultato di un'IA generativa
Prompt	Input in un'IA generativa con le istruzioni al sistema di IA per la generazione dell'output