Il design dell’AI ostacola l’esercizio del diritto all’oblio

di Chiara Temeroli, avvocato, esperta di data protection

Abstract

L'era dell'Intelligenza Artificiale ridefinisce il concetto di memoria digitale, rendendo obsoleto il tradizionale meccanismo della deindicizzazione. Poiché gli LLM integrano le informazioni direttamente nei propri parametri e le rielaborano spesso decontestualizzandole, nemmeno la cancellazione della fonte originaria garantisce l'oblio. La sfida è trovare soluzioni per garantire ad ogni cittadino una reale possibilità di riprogettare il proprio futuro.

Quotidianamente l’Intelligenza Artificiale sta trasformando il modo in cui le informazioni vengono prodotte, conservate ed elaborate. Se ciò, da un lato, ci offre opportunità straordinarie nell'analisi dei dati, dall'altro, mette a dura prova la privacy e il diritto all’oblio, ossia la tutela che impedisce di rimanere "prigionieri del proprio passato".

Questa salvaguardia dell’identità digitale consente di richiedere la rimozione di informazioni personali non più rilevanti, di scarso interesse pubblico o legate a vicende (come processi giudiziari) conclusesi con l'assoluzione, la riabilitazione o a fronte del semplice decorso del tempo. Tuttavia, il diritto all'oblio si scontra oggi con algoritmi capaci di conservare e rielaborare all'infinito dati che dovrebbero essere dimenticati.

Nei motori di ricerca tradizionali, detta protezione ha trovato applicazione pratica soprattutto attraverso la deindicizzazione, ovvero la rimozione dell'associazione tra il nome di una persona e una specifica URL; l'avvento dell'IA generativa ha però complicato radicalmente questo scenario in quanto i sistemi di IA non si limitano a restituire una lista di link, ma assorbono enormi volumi di dati sul web (web scraping) per poi sintetizzarli in risposte apparentemente oggettive.

La deindicizzazione sui motori tradizionali si rivela quindi insufficiente poiché:

- il contenuto rimane visibile sul sito originale;
- l'informazione resta accessibile via link diretto, archivi web o aggregatori secondari;
- i crawler dell'IA possono continuare ad accedere a tali fonti se le rimozioni non vengono aggiornate nei loro registri.

Inoltre, rielaborando notizie datate senza il dovuto contesto, l'IA rischia di generare "allucinazioni", incrociando fatti reali con dettagli errati. Una persona assolta anni fa può così ritrovarsi descritta da una chatbot come "ancora coinvolta in un processo per frode", con gravi danni reputazionali.

Si potrebbe pensare che, eliminando la pagina web originale, il problema possa essere risolto. In realtà, il rischio persiste anche dopo la rimozione del contenuto originale, per le seguenti ragioni fondamentali.

Durante la fase di addestramento, l'algoritmo ha già "digerito" la notizia; l'informazione non si trova più dunque come file isolato, ma è integrata nella struttura matematica della rete neurale. Anche se la fonte sparisce dal web, la memoria dell'IA rimane intatta.

Rimuovere un singolo dato personale da un LLM senza dover riaddestrare l'intero modello da capo (un processo lunghissimo ed economicamente insostenibile) è una sfida tecnica ancora aperta. I filtri di sicurezza applicati dai gestori riescono solo parzialmente a bloccare l'output.

Quando un'IA cerca riscontro sul web in tempo reale e non trova più la fonte originale, non cancella automaticamente il "ricordo" assimilato in precedenza; al contrario, tenta di ricostruire la notizia basandosi su frammenti memorizzati, aumentando drasticamente il rischio di inventare dettagli e generare risposte diffamatorie senza che l'utente possa verificare la fonte.

Prima della cancellazione, la pagina potrebbe essere stata salvata da motori di ricerca (nella versione cache), aggregatori automatici o banche dati indipendenti (come Internet Archive), continuando ad alimentare i dataset di addestramento.

Come noto, nel contesto europeo la salvaguardia dei dati personali trova il suo pilastro nel GDPR, che sancisce il diritto alla cancellazione (Art. 17). Tuttavia, il diritto all'oblio non ha un valore assoluto: deve essere continuamente bilanciato con la libertà di espressione e di informazione, oltre che con l'interesse della collettività alla memoria storica e alla trasparenza dei fatti di rilevanza sociale o politica.

L'avvento dell'Intelligenza Artificiale rende questo equilibrio ancora più fragile. A differenza di un giudice o di un'autorità di garanzia, i crawler e gli algoritmi automatizzati non effettuano valutazioni etiche o giuridiche sul singolo caso, ma elaborano le informazioni come puri dati statistici. Il rischio concreto è che notizie ormai superate continuino ad essere alimentate e diffuse, venendo private della possibilità di perdere rilevanza pubblica.

Rendere effettivo il diritto all'oblio nell'era dell'IA richiede un intervento coordinato su più livelli:

- non appare sufficiente agire sui motori di ricerca ma occorre intervenire direttamente alla fonte, richiedendo al sito originario la cancellazione o l'anonimizzazione dei dati, per poi estendere la richiesta ai gestori dei modelli di IA;
- è fondamentale accelerare la ricerca su tecniche capaci di far "dimenticare" all'algoritmo specifici dati di addestramento senza riaddestrare l'intero sistema;
- è necessario che le aziende sviluppatrici garantiscano la piena tracciabilità dei dataset utilizzati e offrano procedure di reclamo semplici e accessibili per l'esclusione dei dati personali.
- è opportuno promuovere un'adeguata educazione digitale, indispensabile per far comprendere ai cittadini che ogni contenuto immesso in rete lascia un'impronta destinata a essere assimilata dai sistemi automatizzati.

Tutelare l'oblio nell'epoca dell'Intelligenza Artificiale non è un mero esercizio normativo, ma una difesa fondamentale della dignità umana: un presidio che garantisce a ogni individuo la possibilità di rinascere e riprogettarsi, libero dal peso del proprio passato.

AI Design Impedes the Right to Be Forgotten

by Chiara Temeroli, lawyer, data protection expert

Abstract

The era of Artificial Intelligence is redefining the concept of digital memory, making the traditional mechanism of de-indexing obsolete. Since LLMs integrate information directly into their own parameters and often reprocess it decontextualized, not even deleting the original source guarantees oblivion. The challenge is to find solutions to ensure every citizen has a real opportunity to redesign their own future.

Artificial Intelligence is transforming the way information is produced, stored, and processed every day. While this offers us extraordinary opportunities in data analysis, it also challenges privacy and the right to be forgotten—the protection that prevents us from remaining "prisoners of our past."

This digital identity protection allows you to request the removal of personal information that is no longer relevant, of little public interest, or related to events (such as legal proceedings) that ended with acquittal, rehabilitation, or simply the passage of time. However, the right to be forgotten today clashes with algorithms capable of endlessly storing and reprocessing data that should be forgotten.

In traditional search engines, this protection has found practical application primarily through de-indexing, or the removal of the association between a person's name and a specific URL. However, the advent of generative AI has radically complicated this scenario, as AI systems do not simply return a list of links, but absorb enormous volumes of data from the web (web scraping) and then synthesize it into seemingly objective responses.

Deindexing on traditional search engines is therefore insufficient because:

• the content remains visible on the original site;

• the information remains accessible via direct links, web archives, or secondary aggregators;

• AI crawlers can continue to access these sources if the deletions are not updated in their records.

Furthermore, by reprocessing dated news without proper context, AI risks generating "hallucinations," combining real facts with incorrect details. A person acquitted years ago could thus find themselves described by a chatbot as "still involved in a fraud trial," resulting in serious reputational damage.

One might think that deleting the original web page would solve the problem. In reality, the risk persists even after the original content has been removed, for the following key reasons.

During the training phase, the algorithm has already "digested" the news; the information is therefore no longer present as an isolated file, but is integrated into the mathematical structure of the neural network. Even if the source disappears from the web, the AI's memory remains intact.

Removing a single piece of personal data from an LLM without having to retrain the entire model from scratch (a time-consuming and economically unsustainable process) is still an ongoing technical challenge. The security filters applied by operators are only partially effective at blocking the output.

When an AI searches the web for information in real time and no longer finds the original source, it does not automatically delete the previously assimilated "memory"; instead, it attempts to reconstruct the news item based on stored fragments, dramatically increasing the risk of inventing details and generating defamatory responses without the user being able to verify the source.

Before deletion, the page may have been saved by search engines (in cache), automatic aggregators, or independent databases (such as the Internet Archive), continuing to feed the training datasets.

As is well known, in the European context, the protection of personal data is fundamentally based on the GDPR, which enshrines the right to erasure (Art. 17). However, the right to be forgotten is not absolute: it must be continually balanced with freedom of expression and information, as well as with the public interest in historical memory and the transparency of events of social or political relevance.

The advent of Artificial Intelligence makes this balance even more fragile. Unlike a judge or a supervisory authority, crawlers and automated algorithms do not perform ethical or legal assessments on a case-by-case basis, but process information as pure statistical data. The real risk is that outdated news will continue to be fueled and disseminated, depriving it of the possibility of losing public relevance.

Enforcing the right to be forgotten in the AI era requires coordinated intervention on multiple levels:

• It is not sufficient to simply target search engines; it is necessary to intervene directly at the source, requesting the original site to delete or anonymize the data, and then extending the request to AI model managers;

• It is essential to accelerate research into techniques capable of making the algorithm "forget" specific training data without retraining the entire system;

• Developers must ensure full traceability of the datasets used and offer simple and accessible complaints procedures for the exclusion of personal data.

• It is important to promote adequate digital education, which is essential to help citizens understand that every piece of content posted online leaves a mark that will be assimilated by automated systems.

Protecting oblivion in the age of Artificial Intelligence is not merely a regulatory exercise, but a fundamental defense of human dignity: a safeguard that guarantees each individual the opportunity to be reborn and reimagined, free from the burden of their past.