*Cambridge Forum on AI: Law and Governance* (2026), **2**, e12, 1–17
doi:10.1017/cfl.2026.10062

Cambridge Forum

RESEARCH ARTICLE

# Legal remedies for AI-generated reputational harm

Laura Herrerias Castro

Law Department, Pompeu Fabra University, Barcelona, Spain
Email: laura.herrerias@upf.edu

(Received 11 February 2026 UTC; revised 16 May 2026 UTC; accepted 18 June 2026 UTC)

**Abstract**

The widespread use of generative artificial intelligence (GenAI) has increased the risk of reputational harm caused by false or "hallucinated" outputs. This article explores how the European Union legal framework, together with selected Member States' legal systems, addresses liability and remedies for such harms, with particular attention to defamation, product liability, noncontractual liability and data protection. Then, the article analyzes the remedies available to injured parties, focusing on claims for damages, injunctive relief and the exercise of data protection rights. While national legal systems generally allow compensation for nonpecuniary damage resulting from violations of personality rights, the amounts awarded are likely to be minimal due to the limited dissemination of AI-generated content. Consequently, the article argues that prohibitory injunctions aimed at preventing the publication of identical or equivalent false information frequently offer a more robust form of protection than financial redress. Finally, operators of GenAI systems must comply with the principles relating to the processing of personal data, particularly the principle of accuracy and ensure the effective exercise of data protection rights; otherwise, data subjects may bring an action for compensation against the data controller under Article 82 of the General Data Protection Regulation.

**Keywords:** generative AI; reputation; right to compensation; injunctive relief; right to rectification; right to be forgotten

## 1. Introduction

Traditionally, users relied on search engines to access information online. Now, they are increasingly employing generative artificial intelligence (GenAI) systems for this purpose. This shift reflects a growing convergence between search technologies and GenAI (Lorente & Gardhouse, 2026, p. 8), as large language models (LLMs) are being embedded within search engines. For example, Microsoft has incorporated GPT-5 into its Bing search engine through Microsoft Copilot, while Google Search has introduced a new search feature known as AI Mode, which uses the Gemini 2.0 model.

LLMs frequently stray far from the truth or "hallucinate" in their quest to be convincing and helpful to users (Metz & Weise, 2025), and are prone to produce oversimplifications of complex topics and responses biased toward certain commonly held beliefs (Wachter et al., 2024, p. 2). Hallucinations can arise from several reasons, including insufficient or low-quality training data, as datasets may contain false or outdated information (Huang et al., 2024, pp. 7–10; Zhang et al., 2023, pp. 4–5). In addition, training and evaluation methods often reward guessing over acknowledging uncertainty (Kalai et al., 2025, p. 4). Another source of error occurs when users include incorrect facts in their

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Downloaded from https://www.cambridge.org/core. 18 Nov 2026 at 02:45:14, subject to the Cambridge Core terms of use.

prompts. Because chatbots are designed to produce contextually appropriate responses, they may end up playing along with these inaccuracies rather than correcting them (Jones, 2025).[1]

A user who finds out that a GenAI system generates false information about him or her could reasonably assume that others conducting similar searches could find the same information. Empirical research has shown that users often struggle to distinguish between information generated by humans and that produced by GenAI in terms of credibility and reliability, even when those systems include disclaimers acknowledging the possibility of errors (Huschens et al., 2023, p. 11).

In the GPT-4 System Card, OpenAI expressly warns that:

> Despite GPT-4's capabilities, it maintains a tendency to make up facts, to double-down on incorrect information, and to perform tasks incorrectly. Further, it often exhibits these tendencies in ways that are more convincing and believable than earlier GPT models (e.g., due to authoritative tone or to being presented in the context of highly detailed information that is accurate), increasing the risk of overreliance. (OpenAI, 2023, p. 59)[2]

In this regard, the Confederation of European Data Protection Organizations has observed that the tone adopted by GenAI chatbots may convey an unwarranted sense of certainty and authority: "When generative AI chatbots are palpably wrong or inaccurate, they are often wrong in a very confident and confusingly definitive manner, an attitude that masks the fact that, as OpenAI, for instance admits, the answer may simply be 'nonsense'" (CEDPO, 2023, p. 5).

The dissemination of false information about a person can severely harm their reputation, as illustrated by the following examples. In 2023, Brian Hood, who was the Mayor of Hepburn Shire Council (Australia), filed a lawsuit against OpenAI after ChatGPT falsely claimed he had served time in prison for bribery. Although the lawsuit was dropped in 2024, Mr Hood succeeded in preventing ChatGPT from providing any information about him (and anyone named Brian Hood). That same year, Jeffery Battle – a US Air Force veteran – filed a defamation claim against Microsoft because Copilot mistakenly associated him with a convicted terrorist named Jeffrey Battle.[3] Finally, Mark Walters – an American radio host and Second Amendment advocate – sued OpenAI for defamation because ChatGPT fabricated a legal complaint accusing him of embezzlement.[4]

This article explores how European Union (EU) legislation addresses reputational harm caused by AI-generated content. The legal framework of reference is EU law, particularly the Digital Services Act (DSA),[5] the new Product Liability Directive (PLD),[6] the Artificial Intelligence Act[7] (AIA) and the General Data Protection Regulation (GDPR).[8] In the absence of harmonized rules on defamation and

[1] Mitigation strategies include refining training datasets, incorporating real-time knowledge bases and enhancing human oversight during model fine-tuning (Ferraris et al., 2025, p. 15).

[2] The problem of hallucinations persists in the current GPT-5.5 model, despite some improvements (OpenAI, 2026a, p. 12).

[3] *Battle v. Microsoft Corporation*, 1:23-cv-01822 (D. Maryland, 7.7.2023).

[4] *Walters v. OpenAI, L.L.C.*, 1:23-cv-03122 (N.D. Ga., 14.7.2023). On 19 May 2025, the Superior Court of Gwinnett County (Georgia, USA) ruled in favor of the defendant for three reasons: (a) a reasonable reader wouldn't have understood the allegations as describing actual facts, (b) Walters failed to demonstrate OpenAI's fault under either the negligence or actual malice standards, and (c) he couldn't demonstrate actual damages.

[5] Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (OJ L 277, 27.10.2022).

[6] Directive (EU) 2024/2853 of the European Parliament and of the Council of 23 October 2024 on liability for defective products and repealing Council Directive 85/374/EEC (OJ L, 2024/2853, 18.11.2024).

[7] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonized 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 (OJ L, 2024/1689, 12.7.2024).

[8] 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 (OJ L 119 4.5.2016).

Downloaded from https://www.cambridge.org/core. 18 Nov 2026 at 02:45:14, subject to the Cambridge Core terms of use.

noncontractual liability, the article also draws on Spanish, French and German legislation and case law.

The article is structured as follows: First, it identifies the legal grounds on which GenAI operators may incur liability,[9] focusing on defamation law (section 2.1), product liability (section 2.2), noncontractual liability (section 2.3) and data protection law (section 2.4). Then, it analyzes the remedies available to injured parties, including claims for damages (section 3.1), injunctive relief (section 3.2), the exercise of the right to rectification (section 3.3) and the right to be forgotten (section 3.4), as well as the right to compensation under Article 82 GDPR (section 3.5). The article concludes with an outlook in section 4.

## 2. Legal grounds for reputational harm caused by GenAI

In principle, GenAI systems do not fall within any of the safe harbors regulated under Articles 4–6 of the DSA. Recital 18 DSA expressly clarifies that those exemptions do not apply "in respect of liability relating to information provided not by the recipient of the service but by the provider of the intermediary service itself." In this regard, most GenAI systems play an active role in content creation, rather than merely storing information provided by users (Hacker et al., 2023, p. 1118). This active role places GenAI systems closer to content providers than to mere intermediaries (Melian Pérez, 2025, p. 210). Even if certain services offered by a GenAI operator were to fall within the scope of Article 6.1 DSA, the Court of Justice of the European Union (CJEU) has made clear that the possible benefit of this safe harbor cannot affect the application of the GDPR regime to that operator.[10]

Having established that, in general, GenAI operators cannot benefit from safe harbor protection, the next step is to identify the legal grounds on which they might be held liable for reputational harm resulting from AI-generated outputs.

### *2.1. Defamation law*

The EU has neither harmonized substantive law on defamation nor the conflict-of-law rules in that field.[11] Consequently, each court applies the law designated as applicable under its national conflict rules. In Spain, Article 7.7 of the Organic Act 1/1982, of 5 May, on the civil protection of the right to honor, to privacy and to one's own image,[12] defines defamation as "the imputation of facts or the manifestation of value judgments by acts or expressions that infringe in any way the dignity of another person, damaging his fame or impinging on his self-esteem." In a similar vein, Article 29 of the French Act on the Freedom of the Press, of 29 July 1881,[13] defines defamation as "any allegation or imputation of a fact which is prejudicial to the honor or reputation of the person or entity to which the fact is attributed."

In these Member States, a key hurdle for plaintiffs in defamation cases involving GenAI systems is establishing that the output constitutes a conscious attribution of fact.[14] This difficulty arises from the nature of these systems, which function as probabilistic, predictive tools without communicative intent or subjective reasoning. In this respect, national case law on the liability of search engine operators for "autocomplete" suggestions may offer a useful point of reference.

[9] References to "GenAI operators" encompass both providers and deployers of AI systems (see Art. 3.8 AIA).

[10] Case C-492/23 (2 December 2025), *Russmedia*, ECLI:EU:C:2025:935, para. 131.

[11] Article 1.2(g) of Regulation (EC) No 864/2007 of the European Parliament and of the Council of 11 July 2007 on the law applicable to noncontractual obligations (Rome II) (OJ L 199, 31.7.2007).

[12] Ley Orgánica 1/1982, de 5 de mayo, de protección civil del derecho al honor, a la intimidad personal y familiar y a la propia imagen (BOE nº 115, 14.05.1982).

[13] Loi du 29 juillet 1881 sur la liberté de la presse.

[14] On the difficulties of applying the English tort of defamation to GenAI operators, see Binns & Edwards (2025).

Downloaded from https://www.cambridge.org/core. 18 Nov 2026 at 02:45:14, subject to the Cambridge Core terms of use.

For instance, in Spain, the judgment of the Court of Appeal of Almería 701/2019, of 22 October, held that a search engine, through its autocomplete function, could not be deemed to have committed an act falling within the scope of Article 7.7 of the Spanish Organic Act 1/1982:

> The autocomplete function operates involuntarily, mechanically; therefore, there is no intentionality. The mechanical process carried out by the autocomplete function is evident: an algorithm generates the search predictions without human intervention, and does so based on objective factors, including the frequency with which other users have searched for a particular word. It does not involve a direct attribution of facts or value judgments; rather, it is users who, by repeatedly entering the name of a company alongside a specific activity or concept, cause these associations to appear in the prediction tool, that is, they link the search to those terms.

Similarly, the French Supreme Court in its judgment 12-17.591, of 19 June 2013, concluded that:

> In order to qualify under the Act of 29 July 1881, the statement must be the result of the expression of a thought or intellectual reasoning belonging to its author; the mere juxtaposition of words resulting solely from a systematic and automated classification process, intended to facilitate internet searches, does not constitute the expression of a thought (…) the automatic and instantaneous display of keywords in the search bar on the screen is devoid of any intrinsic meaning and constitutes a mere search aid for internet users by referring them to websites likely to contain the said keywords.

Finally, in its judgment of 14 May 2013, the German Federal Court of Justice (BGH) held Google liable for associating the plaintiff's name with the terms "Scientology" and "fraud" through its autocomplete function. In this regard, the BGH reasoned that:

> A user seeking information through the defendant's search engine expects the search suggestions displayed after entering a query to bear a substantive connection to that query, or at least considers such a connection possible. (…) The Court of Appeal failed to take this expectation into account when assessing the meaning conveyed by the search suggestions displayed by the defendant's search engine. In the present case, this leads to the conclusion that the search suggestions 'Scientology' and 'fraud,' which automatically appear upon entering the full name of the plaintiff, convey the message that there is a substantive connection between the plaintiff and those negatively connoted terms.

The BGH expressly held that Google has no obligation to proactively filter content and is required to act only upon obtaining knowledge of its illegality. Under German law, entities that facilitate the distribution of illegal content can be held liable as interferers (*Störerhaftung*), if the following conditions are met: (a) the secondary infringer (*Störer*) contributed to the infringement of the protected rights in an adequately causal manner, (b) they had the possibility of preventing the illegal content and (c) they breached a recognizable duty of care which, according to case law, generally arises once they have knowledge of the specific infringement (Nordemann, 2011, pp. 39-40).[15]

Although autocomplete works differently from a GenAI system, both involve outputs generated through automated processes without any editorial control by the operator.[16] LLMs cannot think or reason, they predict the next token of a string of text based on patterns observed in their training data (Hacker et al., 2025, p. 2). Therefore, claims based on human actions or omissions – particularly an operator's failure to block specific illegal outputs or to comply with data protection obligations – are

[15]See judgments of the BGH of 2 June 2022, I ZR 140/15 (*YouTube II*); 1 March 2016, VI ZR 34/15 (*Jameda.de II*); and 12 July 2012, I ZR 18/11 (*Alone in the Dark*).

[16]It should be noted that developers of LLMs exercise some editorial control, at least insofar as they determine the datasets used to train the model and shape its behavior through fine-tuning.

Downloaded from https://www.cambridge.org/core. 18 Nov 2026 at 02:45:14, subject to the Cambridge Core terms of use.

more likely to succeed than those based on the fact that the operator itself has made a false statement about someone.

### 2.2. Product liability

The new PLD lays down common rules on the liability of economic operators for damage suffered by natural persons and caused by defective products placed on the market or put into service after 9 December 2026. One of the aims of the PLD is to clarify that software – including AI systems – qualifies as a product irrespective of the mode of its supply or usage (Article 4.1 PLD).

When assessing the defectiveness of a GenAI model or system, it is necessary to consider the impact of any capacity the product has to learn or acquire new features after it is placed on the market or put into service (Article 7.2(c) PLD). According to Recital 32 PLD, a manufacturer that designs a product with the ability to develop unexpected behavior should remain liable for behavior that causes harm, unless the manufacturer can rely on the development risk defense (Article 11.1(e) PLD).

However, Article 6 PLD only covers the following types of compensable damage: (a) death or personal injury, including medically recognized damage to psychological health; (b) damage to, or destruction of, any property, with some exceptions; and (c) destruction or corruption of data that are not used for professional purposes. Types of damage other than those provided for in the PLD, such as pure economic loss, privacy infringements or discrimination, should not by themselves trigger liability under the Directive (Recital 24 PLD). Since the PLD excludes the protection of other interests not expressly mentioned, it cannot be applied to reputational harm.

### 2.3. Noncontractual liability

With the exception of the PLD, the EU has not adopted any liability rules specifically addressing AI systems.[17] In the absence of harmonized rules, noncontractual liability will continue to be governed by the national rules of each Member State. This inevitably generates legal uncertainty and fragmentation (Li & Heine, 2026, p. 11).

Within Member States' legal systems, fault-based liability rules provide the backbone of the law of torts. GenAI operators are aware that their systems sometimes generate false or inaccurate outputs. However, foreseeability cannot be assessed in abstract terms; it must be evaluated in relation to whether the party who caused the harm could have reasonably foreseen the specific outcome of their conduct. If the harm is unforeseeable, there cannot be fault in the sense of a breach of a duty of care that was in place to prevent that harm from materializing (Noto la Diega & Bezerra, 2024, p. 20).

Where GenAI operators become aware of a specific illegal output – typically following a user notification[18] – and fail to take reasonable measures to block it, liability should arise. This would mean subjecting GenAI operators to a liability regime similar to that applicable to online platforms for hosting defamatory content (Herrerías Castro, 2026, pp. 333–334). The required standard of conduct must be assessed in light of the state of the art in content-moderation technologies. Accordingly, operators should not be required to ensure perfect suppression: taking reasonable steps to block certain output should be enough to avoid liability (Volokh, 2023, p. 519).

At the same time, liability may arise independently of any prior notification where the defamatory output results from a failure to adopt adequate preventive measures during the design, training, or deployment of the AI system. This approach is reflected in the recent judgment of the Landgericht Kiel

[17] On 28 September 2022, the EU Commission published a proposal for a Directive on adapting noncontractual civil liability rules to artificial intelligence (Brussels, 28.9.2022, COM(2022) 496 final, 2022/0303). Nonetheless, on 11 February 2025, the Commission announced the withdrawal of the proposal due to the absence of a foreseeable agreement among the Member States (European Commission, 2025, p. 26).

[18] As established in Article 16.3 DSA, notices shall be considered to give rise to knowledge of the specific item of information concerned where they allow a diligent operator to identify the illegality of information without a detailed legal examination.

Downloaded from https://www.cambridge.org/core. 18 Nov 2026 at 02:45:14, subject to the Cambridge Core terms of use.

of 29 February 2024 (6 O 151/23). In that case, the defendant operated a website providing economic information on German companies. The service relied on AI tools that processed data from official sources and automatically generated structured summaries for users. Due to an allocation error, the claimant, a medium-sized family business, was incorrectly identified as being subject to deletion for lack of assets.

The court held that the defendant qualified as a direct interferer (*unmittelbare Störerin*) because it voluntarily uses its own software to respond to search queries. The operator could not avoid responsibility by invoking the automated nature of the process, since it had consciously chosen to employ an AI system that was inadequately programmed. The court further emphasized that an AI operator may incur liability where, from the perspective of a reasonable average user, it appears to adopt the generated content as its own and thereby assumes editorial responsibility for it.

This reasoning may have important implications for both AI providers and deployers. For example, a provider that trains a model on low-quality or unreliable datasets – such as unverified online forums – may be regarded as having failed to adopt adequate safeguards against foreseeable hallucinations. Similarly, a deployer that knowingly integrates a highly hallucinatory model into a service intended to provide factual information could arguably be found to have breached its duty of care.

Finally, liability rules are not conceived in isolation from product safety regulation.[19] Although GenAI systems causing reputational harm will not generally qualify as "high-risk AI systems," the obligations imposed by Chapter III AIA may nevertheless provide valuable guidance for defining the applicable standard of care. In particular, obligations relating to data governance (Article 10), transparency (Article 13), human oversight (Article 14), and accuracy, robustness and cybersecurity (Article 15) may serve as benchmarks for assessing whether an AI provider or deployer acted with due care in the design and deployment of an AI system.

### 2.4. Data protection

Personal data processing involves multiple entities and various purposes that depend on the stage of the LLM life cycle. These stages can be summarized in two phases: the development phase, which includes all stages before the deployment of the LLM (e.g., collection of data, preprocessing and training), and the deployment phase, which covers all stages related to the use of the LLM (EDPB, 2024a, p. 11).[20] During the deployment phase, personal data may be processed both as input (prompts including personal data) and as output (inferences including personal data) of the AI system, but also due to model memorization of training data (reproduction of personal data) (EDPS, 2025a, p. 12).

Article 5.1(d) GDPR establishes that personal data must be accurate. This principle imposes two obligations on data controllers: first, to ensure that personal data are kept up to date; and second, to ensure that inaccurate personal data are erased or rectified without delay. Recital 39 GDPR further clarifies that these obligations are subject to a proportionality standard, providing that: "every reasonable step should be taken to ensure that personal data which are inaccurate are rectified or deleted." Accordingly, data controllers must ensure data accuracy throughout the life cycle of a GenAI system. This obligation may be fulfilled by using high-quality training data and incorporating human oversight into the AI decision-making process (EDPS, 2025b, pp. 26–27). Noncompliance with the principle of accuracy in the area of GenAI has already led to a complaint to the Austrian data protection authority.[21]

[19] In this regard, Article 7 of the new PLD underscores the relevance of compliance with product safety requirements when assessing whether a product is defective.

[20] On the processing of personal data in the GenAI data life cycle, see Rubí Puig (2024, pp. 957–967); Nunez Duffourc et al. (2024, pp. 244–249). On the legal basis for data processing across the life cycle of a GenAI system, see Ruschemeier (2025, pp. 6–10).

[21] Information available in: https://noyb.eu/sites/default/files/2024-04/OpenAI%20Complaint_EN_redacted.pdf [12/05/2026].

Downloaded from https://www.cambridge.org/core. 18 Nov 2026 at 02:45:14, subject to the Cambridge Core terms of use.

Data controllers must also facilitate the exercise of data subject rights and inform data subjects of the existence of their rights to request the rectification or erasure of personal data (Articles 13.2(b) and 14.2(c) GDPR). In this regard, Google's Gemini Apps Privacy Hub provides that:

> LLM experiences (Gemini Apps included) can hallucinate and present inaccurate information as factual. Under certain privacy laws, including the General Data Protection Regulation in the EU, you may have the right to: 1. Object to the processing of your personal data, or 2. Ask for inaccurate personal data in Gemini Apps' responses to be corrected. To exercise these rights, you can create a request in our Help Center. You can also create a request directly in Gemini Apps. Below the Gemini app's response, select More: Report legal issue.[22]

Similarly, OpenAI's privacy policy includes the following disclaimer:

> A note about accuracy: Services like ChatGPT generate responses by reading a user's request and, in response, predicting the words most likely to appear next. In some cases, the words most likely to appear next may not be the most factually accurate. For this reason, you should not rely on the factual accuracy of output from our models. If you notice that ChatGPT output contains factually inaccurate information about you and you would like to request a correction or removal of the information, you can submit these requests through privacy.openai.com or to dsar@openai.com, and we will consider your request based on applicable law and the technical capabilities of our models.[23]

Whereas in early 2023 only very basic mechanisms – if any – for exercising data protection rights were available, such as a simple email address; by the end of that year, most GenAI system operators had implemented more detailed and accessible complaint mechanisms for users (Edwards et al., 2025, p. 8). Although GenAI operators formally allow the exercise of the rights to rectification and to be forgotten, several difficulties can hinder their effective implementation. These challenges are examined in the next section.

## 3. Remedies

Remedies for reputational harm caused by AI-generated content are primarily rooted in noncontractual liability. This route of liability is particularly relevant where the victim is a legal entity and thus falls outside the personal scope of the GDPR.

Depending on the Member State, the applicable legal basis may lie either in the general clause on extra-contractual liability or in specific legal regimes. For example, in Germany, claims would typically be brought under § 823(1) of the *Bürgerliches Gesetzbuch* (BGB), which safeguards protected interests, including general personality rights (*Allgemeines Persönlichkeitsrecht*). In contrast, in Spain, the civil protection of the rights to honor, to privacy and to one's own image is governed by the Organic Act 1/1982, of 5 May.

In addition, data subjects may also invoke their rights under Articles 15–22 GDPR, in particular the rights to rectification and to be forgotten, where AI-generated content involves the processing of inaccurate or unlawful personal data.

### *3.1. Damages*

Violations of personality rights may be redressed through claims for pecuniary and nonpecuniary loss. Member States' national legal systems generally compensate for nonpecuniary loss, although the extent and conditions of compensation vary from country to country (Rogers, 2001, pp. 280–281).

[22] Google (2026, May 5).
[23] OpenAI, (2026b, April 1).

Downloaded from https://www.cambridge.org/core. 18 Nov 2026 at 02:45:14, subject to the Cambridge Core terms of use.

Some legal systems do not distinguish between pecuniary and nonpecuniary damages for the purposes of compensation (e.g., France, Belgium and Spain). Others, by contrast, exclude compensation for nonpecuniary damage as a matter of principle and allow it only in a limited number of expressly defined circumstances.[24] This may be done by way of a general provision or by statutory provisions (Zimmermann, 2011, pp. 706–708).

Infringements of personality rights constitute an important case-scenario for awarding nonpecuniary damages. Where a general right of personality is recognized, in general only serious infringements justify an award of nonpecuniary damages (Warzilek, 2005, p. 645). The opposite approach is adopted by provisions like Article 9.3 of the Spanish Organic Act 1/1982, which unrebuttably presumes nonpecuniary losses.[25] The main difficulties associated with nonpecuniary damages lie in their assessment and quantification. Unlike pecuniary damages, which can be measured by reference to objective market values, nonpecuniary damages lack an objective benchmark. Therefore, their quantification depends on subjective evaluation, leading to significant variability in the compensation awarded.

According to Article 10:301 (2) PETL, in the assessment of nonpecuniary damages, "all circumstances of the case, including the gravity, duration and consequences of the grievance, have to be taken into account." Similarly, Article 9.3 of the Spanish Organic Act 1/1982 provides that the amount of compensation for nonpecuniary loss shall be assessed based on the circumstances of the case and the seriousness of the harm, and to this end, special consideration shall be given to the dissemination or audience of the media through which the harm occurred. On 13 January 2026, the Spanish Government announced a bill to adapt the Organic Act 1/1982 to the digital environment.[26] One of its innovations is the introduction of additional criteria for assessing nonpecuniary damages, including: the violation of one or more fundamental rights, the gravity of the expressions used and their social repercussions, the duration of the infringement and the distress or suffering caused by efforts to bring about its cessation.

In online defamation cases, Spanish courts have awarded nonpecuniary damages against hosting service providers in amounts ranging from €1,000 to €18,000, depending on the abovementioned criteria.[27] Compensation tends to be significantly higher where the violation of the right to honor occurs through traditional media services. Judgment 390/2017, of 17 November, of the Court of Appeal of Valladolid justifies this differentiation for the following reason: "The internet indeed multiplies the number of potential recipients of messages, but it is also true that the purpose of such messages is not usually general, or at least not as general as their dissemination through the media (newspapers, radio, television)." This observation is particularly pertinent in the context of searches conducted through GenAI systems, where access to information is typically initiated by specific user queries.

[24]In this regard, Article 10:301 (1) of the Principles of European Tort Law (PETL) establishes that: "Considering the scope of its protection (article 2:102), the violation of an interest may justify compensation of non-pecuniary damage. This is the case in particular where the victim has suffered personal injury; or injury to human dignity, liberty, or other personality rights". On the protected interests under Article 2:102 PETL, see Infantino (2023).

[25]See judgment of the Spanish Supreme Court 742/2016, of 21 December 2016 (ECLI:ES:TS:2016:5659), according to which Article 9.3 establishes a presumption *iuris et de iure* of the existence of nonpecuniary damage when there has been an illegitimate invasion of the right to honor.

[26]The text can be consulted at: https://www.mpr.gob.es/servicios/participacion/audienciapublica/Paginas/2026/2025-1206%20APLO%20Honor/anteproyecto-de-ley-org-nica-de-protecci-n-civil-d.aspx [12/05/2026].

[27]Judgments of the Court of Appeal of Santa Cruz de Tenerife, sec. 3, 1/2022, 13 January, ECLI:ES:APTF:2022:97 (compensation of €5,000); Madrid, sec. 12, 47/2015, 4 February, ECLI:ES:APM:2015:4445 (compensation of €10,000); Málaga, sec. 4, 540/2011, 24 October, ECLI:ES:APMA:2011:1605 (compensation of €10,000); Madrid, sec. 11, 221/2011, 31 March, ECLI:ES:APM:2011:2467 (compensation of €9,000); Barcelona, sec. 14, 707/2010, 29 November, ECLI:ES:APB:2010:8805 (compensation of €12,000); Badajoz, sec. 3, 280/2010, 17 September, ECLI:ES:APBA:2010:871 (compensation of €2,000); Madrid, sec. 13, 420/2008, 22 September, ECLI:ES:APM:2008:18214 (compensation of €6,000); Islas Baleares, sec. 3, 65/2007, 22 February, ECLI:ES:APIB:2007:200 (compensation of €6,000); and Madrid, sec. 19, 50/2006, 6 February, ECLI:ES:APM:2006:266 (compensation of €18,000).

Downloaded from https://www.cambridge.org/core. 18 Nov 2026 at 02:45:14, subject to the Cambridge Core terms of use.

In most cases, the amount of compensation will be minimal, given the limited dissemination of the output. While in common law systems, especially in the US,[28] claims for nominal damages are allowed, most Member States consider nominal or symbolic damages to be inadmissible (Magnus, 2001, p. 187). For example, in judgment 232/2020 of 2 June, the Spanish Supreme Court held that an award of €100 for nonpecuniary damage arising from an infringement of the right to honor was merely symbolic and therefore upheld the first-instance award of €2,000. Similarly, the Spanish Supreme Court judgment 416/2014, of 24 July, described compensation of €300 as purely symbolic given the circumstances of the case and adopted the amount set in the first instance, which was €6,000.

### 3.2. Injunctions

In the majority of Member States, claims for removal to undo current invasions of personality rights are widely available (Von Bar, 2000, pp. 146–147). It is also possible to seek a prohibitory injunction to prevent future illegal publications (e.g., Articles 1004.1 BGB, 9.2 of the French Civil Code and 9.2(b) of the Spanish Organic Act 1/1982).[29] Both remedies are complementary, since they are addressed at removing the infringement and preventing it from persisting into the future.

In the context of the liability of online intermediaries for hosting defamatory content, the CJEU held that an injunction ordering a social network to remove information that is identical to, or conveys an equivalent meaning as, content previously declared defamatory does not infringe Article 15.1 of the E-Commerce Directive (now Article 8 DSA):

> for an injunction which is intended to bring an end to an illegal act and to prevent it being repeated, in addition to any further impairment of the interests involved, to be capable of achieving those objectives effectively, that injunction must be able to extend to information, the content of which, whilst essentially conveying the same message, is worded slightly differently, because of the words used or their combination, compared with the information whose content was declared to be illegal.[30]

As regards equivalent or similar content, the CJEU further clarified that:

> it is important that the equivalent information (…) contains specific elements which are properly identified in the injunction, such as the name of the person concerned by the infringement determined previously, the circumstances in which that infringement was determined and equivalent content to that which was declared to be illegal. Differences in the wording of that equivalent content, compared with the content which was declared to be illegal, must not, in any event, be such as to require the host provider concerned to carry out an independent assessment of that content.[31]

Such an injunction would not entail an excessive obligation, as the monitoring and search for information it requires are limited to information containing the elements specified in the injunction, and its defamatory content of an equivalent nature does not require the hosting provider to carry out an independent assessment, as the latter has automated search tools and technologies at its disposal.[32]

[28] *Restatement of the Law, Second, Torts*, § 901.

[29] In the case C-655/23 (4 September 2025), *Quirin Privatbank*, ECLI:EU:C:2025:655, para. 52; the CJEU concluded that the provisions of the GDPR must be interpreted as not providing, for the data subject concerned by the unlawful processing of personal data, in the event that data subject does not request that his or her data be erased, a judicial remedy enabling him or her to obtain, as a preventive measure, an order that the controller refrain from any further unlawful processing in the future. However, those provisions do not prevent Member States from providing for such a remedy in their respective legal systems.

[30] Case C-18/18 (3 October 2019) *Glawischnig-Piesczek*, ECLI:EU:C:2019:821, para. 41.

[31] Case C-18/18 (*Glawischnig-Piesczek*), para. 45. See also case C-401/19 (26 April 2022), *Poland*, ECLI:EU:C:2022:297, para. 90.

[32] Case C-18/18 (*Glawischnig-Piesczek*), para. 46.

Downloaded from https://www.cambridge.org/core. 18 Nov 2026 at 02:45:14, subject to the Cambridge Core terms of use.

Although Article 8 DSA does not apply to operators of GenAI systems, as they do not provide intermediary services, the CJEU has held, in the context of measures adopted to protect rightholders, that it is for the national courts to strike a fair balance between, on the one hand, the protection of rightholders' rights under Article 7 CFEU, and, on the other, the freedom to conduct a business enjoyed by operators under Article 16 CFEU, as well as the freedom of expression and of information, safeguarded for internet users under Article 11 CFEU.[33]

Therefore, in the context of GenAI, courts could issue prohibitory injunctions requiring the prevention not only of the same information but also of equivalent information, provided that such information (a) essentially conveys the same message and (b) can be identified based on specific elements set out in the injunction, without requiring the operator to carry out an independent assessment of the content. For instance, where the information consists of the false assertion that I have served a prison sentence for robbery, the injunction could require the prevention of any output that associates my name with a conviction for robbery. However, if new information were to appear on a GenAI system linking my name to criminal sanctions for different offences, such information – although equally false and potentially harmful – would not convey the same message and, therefore, would not fall within the scope of the injunction, nor would it trigger an obligation on the part of the GenAI operator to prevent it.

In addition to injunctive relief, GenAI systems typically provide mechanisms for reporting illegal outputs. To report content to Google, users must first identify the Google product in which the content appears, including Gemini, and then select the applicable reporting ground from four categories: copyright, court orders, personal data/privacy and others.[34] Similarly, ChatGPT allows users to request the removal of, or restriction of access to, illegal content. The reporting web form requires users to provide their name and email address; the relevant product or feature (such as ChatGPT image generation, ChatGPT Search or Sora); the specific URL, where available; an optional screenshot; the country in which the content is alleged to be illegal; the reason for the report (defamation, discrimination, hate speech, unsafe or prohibited products, misleading commercial information, animal harm or welfare issues or other grounds); and an explanation of why the user believes the content is illegal, including references to the applicable legal provisions.[35]

The European Commission is considering whether ChatGPT's search feature should be classified as a very large online search engine under the DSA.[36] If this classification were to apply, OpenAI would be subject to the due diligence obligations set out in Section 5 of the DSA (Bassini & Palumbo, 2026). These obligations would include conducting annual risk assessments of the design and functioning of ChatGPT, with particular attention to its impact on fundamental rights (Article 34 DSA). OpenAI would also be required to implement reasonable, proportionate and effective risk-mitigation measures, which may include adapting the design, features or functioning of the service, as well as adjusting content-moderation processes (Article 35 DSA).

### 3.3. Right to rectification

Data subjects have the right to obtain from the controller, without undue delay, the rectification of inaccurate personal data relating to them (Article 16 GDPR). This provision must be read in the light of Article 5.1(d) GDPR, which enshrines the principle of accuracy, and of Recital 59 GDPR, which establishes that modalities should be provided for facilitating the exercise of the data subject's rights under that regulation, including mechanisms to request and, if applicable, obtain free of charge, in particular, rectification or erasure of personal data.

[33]Joined Cases C-682/18 and C-683/18 (22 June 2021), *YouTube and Cyando*, ECLI:EU:C:2021:503, para. 138. See also Oruç (2022, pp. 183–188).

[34]Information available in: https://support.google.com/legal/troubleshooter/1114905?hl=en#ts=1115658 [12/05/2026].

[35]Information available in: https://openai.com/form/report-content/ [12/05/2026].

[36]See Jahangir (2025, October 29).

Downloaded from https://www.cambridge.org/core. 18 Nov 2026 at 02:45:14, subject to the Cambridge Core terms of use.

The right to rectification can be invoked regardless of which of the legal bases for processing applies. Nonetheless, this right is especially relevant in situations where the data have not been obtained from the data subject, since such situations entail a higher risk of inaccuracy or incompleteness. In practice, this will often be the case when the processing is based on Article 6.1(f) GDPR (EDPB, 2024b, p. 25).

Article 16 GDPR does not specify which evidence may be required by a controller in order to establish the inaccuracy of the personal data which a data subject seeks to have rectified. According to the CJEU, the data subject may be required to provide relevant and sufficient evidence which, in the light of the circumstances of the case, can reasonably be required of that person in order to establish that those data are inaccurate.[37] It must be emphasized that the data subject does not have to claim or demonstrate a particular interest in the rectification of inaccurate data or that the alleged inaccuracy causes any harm.[38]

Research on the localization and modification of information within LLMs is in its infancy. However, recent studies indicate that factual associations in LLMs correspond to localized computational representations, which can be directly edited or updated without the need to retrain the entire model (Meng et al., 2023, 2022).

### 3.4. Right to be forgotten

The right to be forgotten may be exercised by a data subject through the right of erasure (Article 17 GDPR) or the right to object to the processing (Article 21 GDPR), where the conditions for any of those rights are met (Peguera Poch, 2020, p. 498).

GenAI systems operate in a manner that is functionally comparable to search engines. LLMs are trained on datasets that include a substantial proportion of information scraped from the web,[39] whereas search engines use crawlers to systematically scrape web pages and index the content (Zhang et al., 2024, p. 7). For instance, after generating an answer, Gemini provides links to its underlying sources. ChatGPT, by contrast, tends to distinguish between public and nonpublic figures: in the former case, it generally does not cite sources unless explicitly prompted to do so, whereas in the latter it usually includes references.

The processing of personal data carried out in the context of the activity of a search engine must be distinguished from and is additional to that carried out by publishers of websites (EDPB, 2020, p. 5). Search engine operators are responsible not because personal data appear on a website, but because of the referencing of that page. Although delisting requests do not lead to the complete erasure of personal data, as the data remain available on the source website, the impossibility of finding false or inaccurate information by searching against the data subject's name clearly reduces the public impact of the information (Kranenborg, 2020, p. 479). In the context of GenAI, the exercise of the right to be forgotten entails a right to have one's own private information blocked rather than deleted. Because these systems learn from the data but do not hold the data, the information they generate cannot be made unavailable, but only inaccessible (Floridi, 2023, pp. 3–4).

However, the right to be forgotten cannot be exercised when the processing of personal data is necessary for the exercise of the right to freedom of expression and information (Article 17.3(a) GDPR). While, as a general rule, the data subject's rights under Articles 7 and 8 CFEU prevail over the legitimate interest of internet users in accessing the information in question (Article 11 CFEU), the balance depends on the relevant circumstances of each case. In particular, on the nature of the information and its sensitivity for the data subject's private life, as well as on the public's interest in

[37] Case C-247/23 (13 March 2025), *Deldits*, ECLI:EU:C:2025:172, para. 41.

[38] Opinion of AG Collins, C-247/23 (12 September 2024), *Deldits*, ECLI:EU:C:2024:747, para. 47.

[39] However, due to the probabilistic functioning of LLMs, it is possible that the output does not result from a memorization of training data but from a generation based on correlations learned from other data.

Downloaded from https://www.cambridge.org/core. 18 Nov 2026 at 02:45:14, subject to the Cambridge Core terms of use.

access to that information, which may vary according to the role played by the data subject in public life.[40]

The accuracy of the referenced information also constitutes a relevant factor when assessing whether the conditions for application laid down in Article 17.3(a) GDPR are met. In this regard, the data subject is required to provide evidence that can reasonably be expected to demonstrate the manifest inaccuracy of the information. As the CJEU has established:

> Where the person who has made a request for de-referencing submits relevant and sufficient evidence capable of substantiating his or her request and of establishing the manifest inaccuracy of the information found in the referenced content or, at the very least, of a part – which is not minor in relation to the content as a whole – of that information, the operator of the search engine is required to accede to that request for de-referencing. The same applies where the data subject submits a judicial decision made against the publisher of the website, which is based on the finding that information found in the referenced content – which is not minor in relation to that content as a whole – is, at least prima facie, inaccurate. By contrast, where the inaccuracy of such information found in the referenced content is not obvious, in the light of the evidence provided by the data subject, the operator of the search engine is not required, where there is no such judicial decision, to accede to such a request for de-referencing.[41]

In practice, blocking the output of a GenAI system is challenging and no perfect solution exists (Fosch Villaronga et al., 2018, p. 313). Retraining a model from scratch involves huge computational and economic costs, and raises serious environmental sustainability concerns due to the enormous energy and water consumption (Novelli et al., 2024, p. 7). Consequently, researchers have explored alternative approaches that operate on already trained models. One example is machine unlearning (Rubí Puig & Wolsing, 2025, p. 9), which aims to develop methods for the targeted removal of the effect of training data from the trained model, and the targeted suppression of content in a GenAI system's outputs (Cooper et al., 2025, pp. 4–6).

On the one hand, methods for structural removal aim to reduce the amount of retraining while still ensuring the exclusion of targeted training data. Structural removal approaches are commonly classified into two categories: exact unlearning, which involves some degree of retraining, and approximate unlearning, which does not retrain the model but alters its weights and/or architecture to resemble a model that had not learned from the data in question (Shrishak, 2024, pp. 5–7). In exact unlearning, the goal is to remove the influence of specific data points as if they were never part of the training process. This can be achieved by retraining the model after removing the specific data. One example is the training approach known as Sharded, Isolated, Sliced, and Aggregated, which involves splitting the training process into sub-models based on pre-divided data subsets (EDPS, 2024, pp. 19–20).

However, it is complex to precisely define and localize the unlearning targets, such as the subset of the training set or a knowledge concept that needs to be removed (Liu et al., 2025). Moreover, removing information from a model's parameters does not guarantee that this model can never generate related information. For instance, user prompts can reintroduce information whose effects were previously removed from the model's parameters. Another critical challenge is maintaining a balance where the model unlearns specific information without significantly impacting its overall performance and capabilities (Qu et al., 2025, pp. 54–56).

On the other hand, methods for output suppression include mechanisms such as output filters. Output filtering involves real-time monitoring of a GenAI system's responses to detect and block certain information before it reaches users. These mechanisms can be deployed almost instantly to

[40] Cases C-136/17 (24 September 2019), *GC and Others (De-referencing of sensitive data)*, ECLI:EU:C:2019:773, para. 66; and C-131/12 (13 May 2014), *Google Spain and Google*, ECLI:EU:C:2014:317, para. 81.

[41] Case C-460/20 (8 December 2022), *Google (Déréférencement d'un contenu prétendument inexact)*, ECLI:EU:C:2022:963, paras. 72–73.

Downloaded from https://www.cambridge.org/core. 18 Nov 2026 at 02:45:14, subject to the Cambridge Core terms of use.

steer the model behavior toward a more desirable one without impacting its performance in any significant way.

Nonetheless, output suppression techniques provide only weak unlearning guarantees. The underlying information remains encoded in the model's parameters and can still influence, or reappear in, the system's outputs. Moreover, filters can be reverted by users using carefully crafted prompts (Blanco-Justicia et al., 2025, p. 34). There is also a risk of over-blocking, which may have a chilling effect on freedom of expression, particularly when automated filters err on the side of caution to minimize potential liability. The abovementioned case of Brian Hood and OpenAI illustrates this concern: when asked to provide any information about anyone bearing that name, ChatGPT automatically replies: "I'm unable to produce a response."

In summary, the proportionality of a response to the exercise of the right to be forgotten depends, on the one hand, on the sensitivity of the data and the risks that their continued regurgitation or disclosure pose to the data subject, and, on the other hand, on the infringement of the data controller's freedom to conduct a business, which depends on the practical feasibility and cost of the measures to be taken to comply with the data subject's request (Commission Nationale de l'Informatique et des Libertés (CNIL), 2026).

### 3.5. Right to compensation for infringements of the GDPR

The data subject is entitled to compensation for any damage resulting from the disclosure of their personal data where the data controller, having been requested to do so, fails to rectify or erase the data, even where such rectification or erasure is technically possible. Pursuant to Article 82 GDPR, any person who has suffered material or nonmaterial damage as a result of an infringement of this regulation has the right to receive compensation from the controller or the processor for the damage suffered. It follows that the right to compensation laid down in that provision is subject to three conditions: (a) an infringement of the GDPR; (b) the existence of damage, which cannot be presumed merely on the basis that an infringement took place, even if the provision which has been infringed grants rights to natural persons[42]; and (c) a causal link between the damage and the infringement. The controller's liability is conditional upon the existence of fault, which is presumed unless the controller proves that the event giving rise to the damage is in no way attributable to it.[43]

Not only pecuniary damage but also nonpecuniary damage may give rise to a right to compensation under Article 82 GDPR. Recital 75 GDPR clarifies that risk to the rights and freedoms of natural persons may result from personal data processing which could lead to physical, material or nonmaterial damage, including damage to the reputation. The CJEU has interpreted the notion of nonpecuniary damage broadly, arguing that it may include situations in which the data subject experiences a well-founded fear that some of his or her personal data may be subject to dissemination or misuse by third parties in the future.[44] Thus, the loss of control over personal data, even for a short period of time, may constitute nonpecuniary damage provided that the data subject can show that he or she has actually suffered such damage, however slight, bearing in mind that the mere infringement of the provisions of the GDPR is not sufficient to confer a right to compensation.[45] A mere allegation of fear, with no proven negative consequences, or a purely hypothetical risk of misuse by an unauthorized third party cannot give rise to compensation.[46]

More recently, the CJEU has confirmed that the concept of nonmaterial damage contained in Article 82 GDPR encompasses negative feelings experienced by the data subject as a result of an unauthorized transmission of his or her personal data to a third party, such as fear or annoyance,

[42]Case C-741/21 (11 April 2024), *Juris*, ECLI:EU:C:2024:288, para. 37.
[43]Case C-667/21 (21 December 2023), *Krankenversicherung Nordrhein*, ECLI:EU:C:2023:1022, para. 103.
[44]Case 340/21 (14 December 2023), *Natsionalna agentsia za prihodite*, ECLI:EU:C:2023:986, paras. 79-86.
[45]Case C-590/22 (20 June 2024), *PS (Incorrect address)*, ECLI:EU:C:2024:536, para. 33.
[46]Case C-687/21 (25 January 2024), *MediaMarktSaturn*, ECLI:EU:C:2024:72, para. 68.

Downloaded from https://www.cambridge.org/core. 18 Nov 2026 at 02:45:14, subject to the Cambridge Core terms of use.

which are caused by a loss of control over those data, by a potential misuse of those data or by harm to his or her reputation, provided that the data subject demonstrates that he or she is experiencing such feelings, with their negative consequences, on account of the infringement of that regulation.[47]

Imagine an individual, X, who discovers that a GenAI system, when prompted with the question "Who is X?," produces false information, such as alleging that X has committed a serious criminal offence or attributing to X statements or positions that fundamentally contradict their beliefs or values. For the risk to be considered real rather than merely hypothetical, X would need to demonstrate that the false information is not an isolated output, but is likely to recur, namely that similar prompts would generate the same or comparable inaccurate content. Moreover, X must prove the negative consequences of the fear or annoyance. A problem with this type of immaterial damage arises from the fact that, in the absence of documents or witnesses to corroborate the claimant's pain, the main element of evidence becomes the seriousness of the underlying facts (Rubí Puig, 2024, p. 976).

In this regard, the CJEU has interpreted Article 82.1 GDPR as precluding a national rule or practice which makes compensation for nonpecuniary damage subject to the condition that the damage suffered by the data subject has reached a certain degree of seriousness.[48] Accordingly, this provision does not require the alleged nonpecuniary damage to reach a *de minimis* threshold in order to be compensable.[49] Likewise, since the right to compensation does not have a dissuasive or even a punitive function, the amount of damages awarded cannot be set at a level which exceeds full compensation for that damage.[50] Therefore, where the damage suffered by the data subject is not serious, a national court may award minimal compensation, provided that such compensation fully compensates for the damage suffered.[51]

Finally, the CJEU has clarified that compensation payable under Article 82 GDPR is fully compatible with prohibitory injunctions, since the right to compensation fulfils an exclusively compensatory function, whereas the purpose of a prohibitory injunction imposed on the person responsible for the damage is purely preventive.[52] When the damage caused by an infringement of the GDPR also affects a person's reputation or honor, it might also be possible to seek remedies under the legal systems of each Member State.

## 4. Conclusions

Member States' noncontractual liability rules may offer a viable path for claimants in cases of AI-generated reputational harm, despite the difficulties in proving fault on the part of the GenAI provider or deployer, and establishing the existence of actual damage. Cumulatively or alternatively, injured parties may also invoke the data protection rights under the GDPR and seek compensation pursuant to Article 82, where the conditions set out in this provision are met.

However, in most cases, compensation for damage arising from the generation of false or inaccurate information is likely to be limited. This is because, except in the case of public figures or persons of notable relevance, the potential exposure of such information to other users will generally be restricted and difficult to demonstrate. GenAI systems do not always produce identical outputs in response to the same prompt, which further complicates efforts to establish widespread dissemination or the recurrence of inaccurate content. Consequently, even where the existence of damage is established, the amounts awarded are likely to be modest.

[47]Case C-655/23 (4 September 2025), *Quirin Privatbank*, ECLI:EU:C:2025:655, para. 64.

[48]Case C- 340/21 (*Natsionalna agentsia za prihodite*), para. 78; C-300/21 (4 May 2023), *Österreichische Post*, ECLI:EU:C:2023:370, para. 51.

[49]Case C-456/22 (14 December 2023), *Gemeinde Ummendorf*, ECLI:EU:C:2023:988, para. 18.

[50]Case C-667/21 (*Krankenversicherung Nordrhein*), para. 86.

[51]Cases C-507/23 (4 October 2024), *Patērētāju tiesību aizsardzības centrs*, ECLI:EU:C:2024:854, para. 35; C-182/22 and C-189/22 (20 June 2024), *Scalable Capital*, ECLI:EU:C:2024:531, para. 46.

[52]Case C-655/23 (*Quirin Privatbank*), paras. 82–83.

Downloaded from https://www.cambridge.org/core. 18 Nov 2026 at 02:45:14, subject to the Cambridge Core terms of use.

In this context, preventive remedies are likely to be more effective than purely compensatory ones, notwithstanding the challenges associated with their enforcement. While identifying specific data embedded within an LLM's parameters presents significant technical difficulties, some practical solutions may be employed to comply with prohibitory injunctions or to give effect to data protection rights. These solutions, however, tend to be either costly or technically imperfect. The most robust approach involves retraining the model using machine unlearning techniques, such as exact unlearning. Where retraining is not feasible due to cost or technical constraints, filtering mechanisms may be deployed to mitigate certain outputs without removing the underlying data from the model. Nonetheless, filters offer only weak unlearning guarantees and are usually overinclusive.

More broadly, the technical tools currently available to implement these rights and remedies remain at an early stage of development and do not scale well. For this reason, structural solutions should be preferable. These may include prioritizing the anonymization of training data or, where full anonymization is not feasible, ensuring that the LLM is anonymous after its training, thereby reducing the risk of personal data exposure at the model level.

**Acknowledgements.** I would like to thank Profs Martin Ebers and Mónica Navarro Michel for the opportunity to present the findings of this paper at the AI Liability Conference held on 3 October 2025 at the University of Barcelona. I would also like to thank Prof. Antoni Rubí Puig for his valuable comments on earlier versions of this article.

**Funding statement.** This work was supported by the research project "Digitalización y responsabilidad por producto: análisis de la Directiva (UE) 2024/2853 y de su transposición," funded by the Spanish Ministry of Science, Innovation, and Universities, the European Fund of Regional Development and the Spanish State Research Agency (PID2024-158195OB-I00).

**Competing interests.** The author declares none.

## References

**Bassini, M., & Palumbo, A.** (2026). Beyond intermediaries: Generative AI in search of their legal status. *VerfBlog*. https://doi.org/10.59704/887f92f3bcf01a98

**Binns, R., & Edwards, L.** (2025). ChatGPT tells fibs about me: Are data protection and libel adequate tools to protect reputation in the LLM era? In P. Hacker, B. Mittelstadt, S. Hammer, & A. Engel (Eds.), *The Oxford handbook of the foundations and regulation of generative AI*. Oxford Academic, pp. 1–17. https://doi.org/10.1093/oxfordhb/9780198940272.013.0017

**Blanco-Justicia, A., Jebreel, N., Manzanares-salor, B., Sánchez, D., Domingo-Ferrer, J., Collell, G., & Eeik Tan, K.** (2025). Digital forgetting in large language models: A survey of unlearning methods. *Artificial Intelligence Review*, *58*(90). https://doi.org/10.1007/s10462-024-11078-6

**Commission Nationale de l'Informatique et des Libertés (CNIL).** (2026, January 5). Ensuring and facilitating the exercise of data subjects' rights. https://www.cnil.fr/en/ensuring-and-facilitating-exercise-data-subjects-rights

**Confederation of European Data Protection Organizations (CEDPO)** (2023). *Generative AI: The data protection implications*. CEDPO AI Working Group 16 October 2023. https://cedpo.eu/generative-ai-the-data-protection-implications/

**Cooper, A. F., Choquette-Choo, C. A., Bogen, M., Jagielski, M., Filippova, K., Liu, K., Chouldechova, A., Hayes, J., Huang, Y., Mireshghallah, N., & Shumailov, I.** (2025). Machine unlearning doesn't do what you think: Lessons for generative AI policy, research, and practice. https://doi.org/10.48550/arXiv.2412.06966

**Edwards, L., Szpotakowski, I., Cifrodelli, G., Sangaré, J., & Stewart, J.** (2025). Private ordering, generative AI and the 'platformisation paradigm': What can we learn from comparative analysis of models terms and conditions? *Cambridge Forum on AI: Law and Governance*, *1*, e21–17. https://doi.org/10.1017/cfl.2024.11

**European Commission**. (2025). *Commission work programme 2025*. Moving forward together: A Bolder, Simpler. 11 2.2025, COM(2025) 45 final.

**European Data Protection Board (EDPB)**. (2020). *Guidelines 5/2019 on the criteria of the right to be forgotten in the search engines cases under the GDPR (part 1)*. https://www.edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-52019-criteria-right-be-forgotten-search-engines_en

**European Data Protection Board (EDPB)**. (2024a). *Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models*. https://www.edpb.europa.eu/system/files/2024-12/edpb_opinion_202428_ai-models_en.pdf

**European Data Protection Board (EDPB)**. (2024b). *Guidelines 1/2024 on processing of personal data based on Article 6(1)(f)* GDPR. https://www.edpb.europa.eu/our-work-tools/documents/public-consultations/2024/guidelines-12024-processing-personal-data-based_en

Downloaded from https://www.cambridge.org/core. 18 Nov 2026 at 02:45:14, subject to the Cambridge Core terms of use.

**European Data Protection Supervisor (EDPS)**. (2024). *TechSonar 2025 report*. https://www.edps.europa.eu/data-protection/our-work/publications/reports/2024-11-15-techsonar-report-2025_en

**European Data Protection Supervisor (EDPS)**. (2025a). *Generative AI and the EUDPR. Orientations for ensuring data protection compliance when using Generative AI systems*. https://www.edps.europa.eu/system/files/2025-10/25-10_28_revised_genai_orientations_en.pdf

**European Data Protection Supervisor (EDPS)** (2025b). *Guidance for risk management of artificial intelligence systems*. https://www.edps.europa.eu/data-protection/our-work/publications/guidelines/2025-11-11-guidance-risk-management-artificial-intelligence-systems_en

**Ferraris, A. F., Audrito, D., Di Caro, L., & Poncibò, C.** (2025). The architecture of language: Understanding the mechanics behind LLMs. *Cambridge Forum on AI: Law and Governance*, *1*, e11,1–19. https://doi.org/10.1017/cfl.2024.16

**Floridi, L.** (2023). Machine unlearning: Its nature, scope, and importance for a "Delete Culture." *Philosophy & Technology*, *36*(42), 1–8. https://doi.org/10.1007/s13347-023-00644-5

**Fosch Villaronga, E., Kieseberg, P., & Li, T.** (2018). Humans forget, machines remember: Artificial intelligence and the right to be forgotten. *Computer Law & Security Review*, *34*(2), 304–313. https://doi.org/10.1016/j.clsr.2017.08.007

**Google** (2026, May 5). Gemini apps privacy hub. https://support.google.com/gemini/answer/13594961?hl=en

**Hacker, P., Mittelstadt, B., Hammer, S., & Engel, A.** (2025). Introduction to the foundations and regulation of generative AI. In P. Hacker, A. Engel, S. Hammer, & B. Mittelstadt (Eds.), *The Oxford handbook of the foundations and regulation of generative AI* (Online edition). Oxford Academic, pp. 1–28. https://doi.org/10.1093/oxfordhb/9780198940272.013.0001

**Hacker, P., Engel, A., & Mauer, M.** (2023). Regulating ChatGPT and other large generative AI models. In *FAccT '23: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (pp. 1112–1123). https://doi.org/10.1145/3593013.3594067

**Herrerías Castro, L.** (2026). Liability of generative AI for defamation. *European Review of Private Law*, *34*(2/3), 311–334. https://doi.org/10.54648/erpl2026043

**Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., Chen, Q., Peng, W., Feng, X., Qin, B., & Liu, T.** (2024). A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *ACM Transactions on Information Systems*, https://doi.org/10.1145/3703155

**Huschens, M., Briesch, M., Sobania, D., & Rothlauf, F.** (2023). Do you trust ChatGPT? - Perceived credibility of human and AI-generated content. https://doi.org/10.48550/arXiv.2309.02524

**Infantino, M.** (2023). Protected interests under the principles of European tort law (Art 2:102 PETL) – preserving the past for shaping the future. *Journal of European Tort Law*, *14*(1), 42–60. https://doi.org/10.1515/jetl-2023-0004

**Jahangir, R.** (2025, October 29). *EU weighs regulating OPENAI'S ChatGPT under the DSA*. What Does that Mean? https://www.techpolicy.press/eu-weighs-regulating-openais-chatgpt-under-the-dsa-what-does-that-mean/

**Jones, N.** (2025, January 21). AI hallucinations can't be stopped - but these techniques can limit their damage. *Nature*, *637*, 778–780. https://doi.org/10.1038/d41586-025-00068-5

**Kalai, A. T., Nachum, O., Vempala, S. S., & Zhang, E.** (2025). Why language models hallucinate. https://doi.org/10.48550/arXiv.2509.04664

**Kranenborg, H.** (2020). Article 17. Right to erasure ('right to be forgotten'). In C. Kuner, L. A. Bygrave, & C. Docksey (Eds.), *The EU general data protection regulation (GDPR)*. Oxford University Press, pp. 475–484.

**Li, S., & Heine, K.** (2026). Developing a harm-based approach to understand digital vulnerability in the era of AI: A perspective of the European Union. *Computer Law & Security Review*, 60. https://doi.org/10.1016/j.clsr.2026.106266

**Liu, S., Yao, Y., Jia, J., Casper, S., Baracaldo, N., Hase, P., Yao, Y., Liu, C Y., Xu, X., Li, H., Varshney, K R., Bansal, M., Koyejo, S., Liu, Y.** (2025). Rethinking machine unlearning for large language models. *Nature Machine Intelligence*, *7*, 181–194. https://doi.org/10.1038/s42256-025-00985-0

**Lorente, T., & Gardhouse, K.** (2026). Between search and platform: ChatGPT under the DSA. *Internet Policy Review*, *15*(1). https://doi.org/10.14763/2026.1.2084

**Magnus, U.** (2001). Comparative report on the law of damages. In U. Magnus (Ed.), *Unification of tort law: Damages* (pp. 185–217). Kluwer Law International.

**Melian Pérez, G. E.** (2025). From curators to creators: Navigating regulatory challenges for general purpose generative AI in Europe. *Journal of Intellectual Property, Information Technology and Electronic Commerce Law*, *16*(2), pp. 201–221. https://www.jipitec.eu/jipitec/article/view/437/436.

**Meng, K., Sharma, A. S., Andonian, A., Belinkov, Y., & Bau, D.** (2022). Locating and editing factual associations in GPT. *Advances in Neural Information Processing Systems*, *35*, 17359–17372. https://doi.org/10.48550/arXiv.2210.07229

**Meng, K., Sharma, A. S., Andonian, A., Belinkov, Y., & Bau, D.** (2023). Mass-editing memory in a transformer. https://doi.org/10.48550/arXiv.2210.07229

**Metz, C., & Weise, K.** (2025, January 5). AI is getting more powerful, but its hallucinations are getting worse. *The New York Times*. https://www.nytimes.com/2025/05/05/technology/ai-hallucinations-chatgpt-google.html

Downloaded from https://www.cambridge.org/core. 18 Nov 2026 at 02:45:14, subject to the Cambridge Core terms of use.

**Nordemann, J. B.** (2011). Liability for copyright infringements on the internet: Host providers (Content Providers) - The German approach. *Journal of Intellectual Property, Information Technology and Electronic Commerce Law*, *37*, pp. 37–49. https://www.jipitec.eu/jipitec/article/view/65.

**Noto la Diega, G., & Bezerra, L. C. T.** (2024). Can there be responsible AI without AI liability? Incentivizing generative AI safety through ex-post tort liability under the EU AI liability directive. *International Journal of Law and Information Technology*, *32*. https://doi.org/10.1093/ijlit/eaae021

**Novelli, C., Casolari, F., Hacker, P., Spedicato, G., Floridi, L.** (2024). Generative AI in EU law: Liability, privacy, intellectual property, and cybersecurity. *Computer Law & Security Review*, 55. https://doi.org/10.1016/j.clsr.2024.106066

**Nunez Duffourc, M., Gerke, S., & Kollnig, K.** (2024). Privacy of personal data in the generative AI data lifecycle. *New York University Journal of Intellectual Property and Entertainment Law*, *13*(2), pp. 219–268. https://jipel.law.nyu.edu/privacy-of-personal-data-in-the-generative-ai-data-lifecycle/.

**OpenAI** (2023). GPT-4 technical report. https://doi.org/10.48550/arXiv.2303.08774

**OpenAI**. (2026a, April 23). *GPT-5.5 system card*. https://deploymentsafety.openai.com/gpt-5-5

**OpenAI**. (2026b, April 1). *Europe privacy policy*. https://openai.com/policies/eu-privacy-policy/

**Oruç, T. H.** (2022). The prohibition of general monitoring obligation for video-sharing platforms under article 15 of the E-Commerce directive in light of recent developments: Is it still necessary to maintain it? *Journal of Intellectual Property, Information Technology and Electronic Commerce Law*, *13*(3), pp. 176–199. https://www.jipitec.eu/jipitec/article/view/354.

**Peguera Poch, M.** (2020). The right to be forgotten in the European Union. In G. Frosio (Ed.), *The Oxford handbook of online intermediary liability* (pp. 486–502). Oxford University Press.

**Qu, Y., Ding, M., Sun, N., Thilakarathna, K., Zhu, T., Niyato, D.** (2025). The frontier of data erasure: A survey on machine unlearning for large language models. *Computer*, *58*(1), 45–57. https://doi.org/10.1109/MC.2024.3405397

**Rogers, W. V. H.** (2001). Damages for non-pecuniary loss in a comparative perspective. In *Tort and insurance law* (vol 2). Springer, p. 318.

**Rubí Puig, A.** (2024). Inteligencia artificial generativa y daños por infracciones normativas del derecho de protección de datos personales: Un análisis a partir de la jurisprudencia reciente del TJUE sobre el artículo 82 RGPD. In J. A. Moreno Martínez & P. J. Femenía López(coords.). *Inteligencia artificial y derecho de daños: Cuestiones actuales: Acorde al Reglamento (UE) 2024/1689*. Dykinson, (pp. 949–992).

**Rubí Puig, A., & Wolsing, D.** (2025). *Remedying copyright infringement in generative AI. What role for algorithmic destruction, machine unlearning and damages?* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5465935

**Ruschemeier, H.** (2025). Generative AI and data protection. *Cambridge Forum on AI: Law and Governance*, *1*, e6, 1–16. https://doi.org/10.1017/cfl.2024.2

**Shrishak, K.** (2024). AI-complex algorithms and effective data protection supervision: Effective implementation of data subjects' rights. *European Data Protection Board*. https://www.edpb.europa.eu/system/files/2025-01/d2-ai-effective-implementation-of-data-subjects-rights_en.pdf

**Volokh, E.** (2023). Large libel models? Liability for AI output. *Journal of Free Speech Law*, *3*, pp. 489–558. https://www.journaloffreespeechlaw.org/volokh4.pdf

**Von Bar, C.** (2000). *The common European law of torts* (vol 2). Clarendon Press.

**Wachter, S., Mittelstadt, B., & Russell, C.** (2024). Do large language models have a legal duty to tell the truth? *Royal Society Open Science*, 11. https://doi.org/10.1098/rsos.240197

**Warzilek, A.**(2005). Comparative report. In H. Koziol & A. Warzilek Eds., *The protection of personality rights against invasions by mass media*. Tort and Insurance Law. (vol. 13, pp. 613–657). Springer.

**Zhang, D., Finckenberg-Broman, P., Hoang, T., Pan, S., Xing, Z., Staples, M., & Xu, X.** (2024). Right to be forgotten in the era of large language models: Implications, challenges, and solutions. *AI and Ethics*, *5*(3), 2445–2454. https://doi.org/10.48550/arXiv.2307.03941

**Zhang, Y., Li, Y., Cui, L., Cai, D., Liu, L., Fu, T., Huang, X., Zhao, E., Zhang, Y., Xu, C., Chen, Y., Wang, L., Tuan Luu, A., Bi, W., Shi, F., & Shi, S.** (2023). Siren's song in the AI ocean: A survey on hallucination in large language models. *Computational Linguistics*, *51*(4), 1373–1418. https://doi.org/10.48550/arXiv.2309.01219

**Zimmermann, R.** (2011). Comparative report (Categories 11-13). In B. Winiger, Koziol Helmut, Koch Bernhard A, Zimmermann Reinhard, ***et al.*** (Eds.), *Digest of European tort law: Essential cases on damage* (vol. 2, pp. 706–718). Walter de Gruyter.

**Cite this article:** Herrerias Castro, L. (2026). Legal remedies for AI-generated reputational harm. *Cambridge Forum on AI: Law and Governance*, *2*, e12, 1–17. https://doi.org/10.1017/cfl.2026.10062

Downloaded from https://www.cambridge.org/core. 18 Nov 2026 at 02:45:14, subject to the Cambridge Core terms of use.