# Report

This part of the EDPL hosts reports in which our correspondents keep readers abreast of various national data protection developments in Europe, as well as on the most recent questions in different privacy policy areas. The Reports are organised in cooperation with the Institute of European Media Law (EMR) in Saarbrücken (www.emr-sb.de) of which the Reports Editor Mark D. Cole is Director for Academic Affairs. If you are interested in contributing or would like to comment, please contact him at mark.cole@uni.lu.

## Italy

## Italian DPA Fines OpenAI for GDPR Non-Compliance: The Last Episode of the Garante – OpenAI Saga?

*Pier Giorgio Chiara**

### I. Introduction

On 20 December 2024, the Italian Data Protection Authority (*Garante per la Protezione dei dati personali,* hereinafter 'Garante') published the decision of 2 November 2024[1] containing corrective and sanctioning measures against OpenAI, in relation to the ChatGPT service. The processing was found to be in violation of Articles 5(1)(a), 5(1)(d), 5(2), 6, 8, 12, 13, 24, 25(1), 33, 83(5)(a) of the GDPR. Accordingly, the Garante fined OpenAI €15 million. In addition, the US company will have to carry out a six-month information campaign. This report presents and discusses the decision of the Garante in light of the long-standing investigation of OpenAI, which started in March 2023, by addressing the manifold legal challenges, in terms of data protection, brought about by Large Language Models (LLMs), and specifically ChatGPT. In particular, taking into account the EDPB Opinion 28/2024,[2] the report sheds light on the issue of identifying the appropriate legal basis for processing personal data for training LLMs as well as how to meaningfully implement the principle of transparency and the related information obligations toward users when providing generative AI services. Moreover, the report contributes to expanding the literature on the inner workings of LLMs in general, and ChatGPT in particular. Finally, it may offer valuable insight into the likely next 'battlefield' for the Italian DPA and the other EU DPAs alike, which concerns the limitation on the processing of Italian users' personal data by Hangzhou DeepSeek Artificial Intelligence and Beijing DeepSeek Artificial Intelligence, the Chinese companies responsible for the 'DeepSeek' chatbot service, which was ordered on 30 January 2025.[3]

DOI: 10.21552/edpl/2025/1/17

* Pier Giorgio Chiara is Junior Assistant Professor at the University of Bologna, Department of Legal Studies and CIRSFID-Alma AI Research Centre. For correspondence: <piergiorgio.chiara2@unibo.it>.

1 Garante per la protezione dei dati personali, Decision of 2 November 2024 [10085455].

2 EDPB, 'Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models', adopted on 17 December 2024.

3 Garante per la protezione dei dati personali, 'Artificial Intelligence: The Italian Data Protection Authority blocks DeepSeek' (Press release, 30 January 2025) <https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/10097450#english> accessed 31 March 2025.

### II. Background: The Temporary Ban and the Investigation

The decision originates from an *ex officio* investigation initiated following the publication of press reports regarding a data breach that occurred on 20 March 2023. Due to a bug, the ChatGPT service users saw the chat history titles of other users instead of

their own. OpenAI publicly confirmed the incident. Therefore, on 30 March 2023, the President of the Garante imposed an immediate temporary limitation on the processing of Italian users' data by OpenAI.[4]

Interestingly, in the first pages of the late 2024 decision, the Garante summarises the 'ChatGPT controversy' highlights from the beginning and elucidates on the various concerns underpinning the investigation regarding potential violations of the GDPR. The starting point of this journey is the temporary ban, that has been covered in a prior issue of this journal.[5] The Garante lifted that measure provided that OpenAI implemented appropriate measures to ensure that the processing of personal data within the ChatGPT service complied with the GDPR. Specifically, OpenAI was required to: 1) publish a privacy policy to inform data subjects – not only the users of the service – about the data processing activities for the purpose of training the algorithms, the methods of processing as well as the logic underlying the service and the rights to which they are entitled; 2) set up a mechanism for meaningfully implementing the right to restriction of processing as well as, 3) the right to rectification and erasure; 4) signal the privacy policy in a way that could be read before registering to the service; 5) change the legal basis of contractual obligation for processing to train the algorithms; 6) set up a mechanism for meaningfully implementing the right to restriction of processing if the choice of the legal basis were to be legitimate interest; 7) request to Italian users to pass an 'age gate' to exclude minors; 8) submit to the Garante a plan for the adoption of age verification tools to prevent access to the service by users under the age of 13; 9) promote a non-promotional informational campaign across all major Italian mass media channels, subject to prior agreement with the Garante with regard to the content.

OpenAI then notified the Garante that it had complied with the requests.[6] However, the Garante strongly opposed the media campaign because it was conducted without prior notification and agreement with the Authority and invited OpenAI to submit a revised communication campaign plan no later than 19 May 2023.[7] After a follow-up exchange between the parties, the Garante considered that the proposed informational activities were below expectations and unlikely to reach the intended audience effectively.[8]

Alongside the management of the enforcement measure, the Garante obtained the necessary elements for the investigation *via* two requests for information, pursuant to Article 58(1)(e) GDPR. On 4 April 2023, the Garante asked information regarding a) the functioning of ChatGPT; b) the information provided to data subjects regarding data processing; c) the measures adopted to prevent access to the service by users under the age of 13; d) the data breach of 20 March. On 6 October 2023, an integrative request concerned the age verification system.

With regard to the functioning of ChatGPT, OpenAI stated that its model was trained using three primary sources: i) publicly available internet data, ii) licensed third-party data, and iii) user or 'trainer-provided' data.[9] Related to that, OpenAI argued to have implemented technical and organisational measures with a view to minimising impacts on data subjects' rights, such as limiting personal data in training, excluding sites with large volumes of personal data, avoiding dark web sources, using Azure Cognitive Services to remove personal information and instructing external collaborators to exclude personal data from fine-tuning datasets. Furthermore, OpenAI identified legitimate interest under Article 6(1)(f) of the GDPR as the legal basis for data processing in training, assessing thereby the necessity of the processing through a Legitimate Interest Assessment (LIA). Finally, it gave users the option to opt out of using interaction data for training and permanently delete their accounts. Moreover, OpenAI implemented a removal mechanism on its website, linked directly to the privacy policy, allowing users to object to their data being processed for model training and

4 See PG Chiara, 'Italian DPA v. OpenAI's ChatGPT: The Reasons Behind the Investigation and the Temporary Limitation to Processing' (2023) 9(1) European Data Protection Law Review 68 –72.

5 Ibid; Garante per la protezione dei dati personali, Decision of 11 April 2023 [9874702].

6 Note of 28 April 2023 (protocol no. 69713/23); note of 15 May 2023 (protocol no. 78218/23).

7 Note of 18 May 2023 (protocol no. 79806/23).

8 Note of 22 June 2023 (protocol no. 97898/23).

9 In a subsequent note, OpenAI clarified that 99% of pre-training data comes from public sources such as Common Crawl, whereas the remaining 1% is data licensed from third parties. However, neither Common Crawl nor any third parties participate in preparing the datasets for training.

planned an informational campaign both online and in two major Italian newspapers. Regarding the data breach, OpenAI published a post on its website and notified 440 affected Italian users via email: importantly, the company confirmed in a note that compromised data included names, emails, billing details, and partial credit card information of ChatGPT Plus subscribers. The company identified the bug in an open-source library and resolved it on the same day. Finally, OpenAI partnered up with a company, ie Yoti Ltd., to implement a three-layered age verification mechanism, consisting of i) an age estimation system based on a 'selfie' of the user; ii) an ID scan; and iii) a credit card. In this respect, OpenAI produced a Data Protection Impact Assessment (DPIA) for the age verification system, the contract with Yoti as data processor, a whitepaper on age verification, and the updated privacy policy.

Although OpenAI had substantively responded to all the Garante's information requests, on 26 January 2024 the Italian Supervisory Authority notified OpenAI of the initiation of proceedings for the adoption of corrective and sanctioning measures, alleging the violation of Articles 5, 6, 8, 12, 13, 24, 25(1), 33, and 83(5)(d) of the GDPR with regard to the personal data processing carried out by OpenAI via the service ChatGPT of 30 March 2023.[10]

On 11 April 2024, OpenAI was audited by the Garante, reaffirming its willingness to collaborate, as demonstrated by ongoing dialogue that resulted in a letter of commitment and improvements to its models, including in terms of data protection. Regarding transparency, OpenAI stated that it had published extensive information about its systems and that, at the launch of ChatGPT in November 2022, it had not anticipated the high number of users nor targeted Italy, with the service initially available only in English. The company also emphasised that its revenues had been reinvested in improving the service, including privacy aspects, requesting that this be considered in any potential sanctions. Finally, it rejected the criticisms regarding its media campaign, considering them a misunderstanding.

## III. The Decision

To begin, the Garante addressed the issues of jurisdiction and competence. Until 15 February 2024, OpenAI was established in California and had not designated an establishment within the territory of the European Union. Considering Article 3 GDPR, also in light of EDPB Guidelines 3/2018[11] and the relevant CJEU case law,[12] the Garante deemed itself competent to assess, with regard to Italian territory, the compliance of OpenAI's personal data processing with the GDPR up until that date and to exercise the powers granted to it under Article 58 GDPR. The cooperation mechanism, under which enforcement powers are vested in the Member State's supervisory authority where the main or single establishment is located ('one-stop-shop' mechanism), does not apply in this case. Thus, concerning the change in circumstances related to the main or single establishment, the EDPB has clarified that the competence of a supervisory authority —– in this case, the Garante—ceases only when such a change in circumstances becomes effective, and any pending administrative proceedings must be transferred to the supervisory authority of the State where the main establishment is located, here, the Irish DPA.[13] However, this circumstance does not retrospectively invalidate the legal basis of the authority's initial actions.[14] Therefore, the administrative procedure at stake could not be transferred to the Irish DPA, as some of the contested violations occurred before 15 February 2024.

### 1. The Data Breach: Article 33 GDPR

The Garante contested the violation of Article 33 GDPR since it had not received a notification of the data breach that occurred on 20 March 2023, despite the involvement of Italian data subjects and its jurisdiction over the matter. Within the 72-hour timeframe required by Article 33(1) GDPR, OpenAI had notified the Irish supervisory authority of the incident, assuming it would share the information with other authorities, including the Garante. Although

10 Note of 26 January 2024 (protocol no. 10531/24).

11 EDPB, 'Guidelines 3/2018 on the territorial scope of the GDPR (Article 3)'.

12 Joined Cases C-585/08 and C-144/09 *Peter Pammer v Reederei Karl Schlüter GmbH & Co. KG* and *Hotel Alpenhof GesmbH v Oliver Heller* [2010] ECLI:EU:C:2010:740.

13 EDPB, 'Opinion 8/2019 on the competence of a supervisory authority in case of a change in circumstances relating to the main or single establishment', 8.

14 Ibid.

the company had assessed that the incident did not pose a high risk, it had voluntarily published a post on its website to alert all potentially affected users. However, the Garante did not accept these arguments, as the 'one-stop-shop' mechanism was not applicable when the incident occurred. The company should have notified the data breach under Article 33 GDPR to all European data protection authorities whose data subjects were affected by the breach.

## 2. Failure to Identify a Lawful basis for the Processing: Articles 5(2) and 6 GDPR

The Garante contested the violation of Articles 5(2) and 6 GDPR as OpenAI failed to demonstrate, in line with the accountability principle, that (i) it had clearly identified before the start of the processing activity a legal basis for the processing of personal data to train the model underlying ChatGPT (ie, GPT), which has been publicly available since 30 November 2022 – when the GDPR became applicable to OpenAI; and, (ii) it had specified the legal basis in the information provided to data subjects in accordance with Articles 13 and 14 GDPR.[15]

Although OpenAI claimed to have based the processing operations related to the provision of the service to users on Article 6(1)(b) GDPR and the processing operations related to algorithm training on Article 6(1)(f) GDPR, the documentation provided (ie, a copy of the DPIA and the LIA), could not prove that the adequacy assessment and the identification of the legal basis had been formalised as of 30 November 2022, as required by the accountability principle under Article 5(2) and Article 6 GDPR. Furthermore, OpenAI failed to explicitly state in the privacy notice that the processing was based on Article 6(1)(f) GDPR and to specify the legitimate interests pursued. This omission affected data subjects' ability to exercise their right to object under Article 21 GDPR, breaching the accountability principle.

Finally, regarding the assessment of the legitimacy of the decision made after the start of processing to rely on legitimate interest for processing the personal data of users and non-users for algorithm training and service operation, the Garante transmitted the records of the proceeding to the Irish supervisory authority, in accordance with the 'one-stop-shop' mechanism.

## 3. OpenAI's Privacy Policy: Articles 5(1)(a), 12, 13 GDPR

Concerning the violation of Articles 5(1)(a), 12, and 13 GDPR, the Garante held that, as of 30 March 2023, OpenAI's privacy policy was available only in English and was not easily accessible on the website. Specifically, the link provided in the registration flow was positioned in a way so that users could not read it before entering their data to create an account. Furthermore, the information provided at the time by the data controller referred exclusively to the users' data processed by OpenAI for the use of ChatGPT. Yet, users and non-users were not provided with any information regarding the processing of personal data to train the LLMs.

With regard to the information relating to the processing of users' chats for developing and improving GPT models, the Garante observed that the privacy policy lacked clarity, particularly because the generic purpose of service improvement made it difficult to connect with the specific and innovative purpose of training generative artificial intelligence models. On the other hand, despite OpenAI arguing it published, since 2019, several research and technical documents, articles, and posts that detailed OpenAI's use of publicly available internet data for training its language models, the Garante considered that the documentation provided by OpenAI could not fulfill the transparency obligation under the GDPR, as the company did not clarify why data subjects would have been expected to access such documents.

Interestingly, the Garante considered that OpenAI breached not only the transparency obligations pursuant to Articles 12 and 13 GDPR, but also the transparency principle itself (Article 5(1)(a) GDPR) due to the critical lack of information provided both to users and non-users vis-à-vis the processing of their personal data.

## 4. Technical and Organisational Measures, Data Protection by Design and by Default: Articles 24, 25(1) GDPR

At least until 30 March 2023, OpenAI did not foresee any age verification system for the registration

15 See EDPB, 'Guidelines 2/2019 on the processing of personal data under Article 6(1)(b) GDPR in the context of the provision of online services to data subjects'.

phase to its service ChatGPT. However, ChatGPT's terms and conditions considered minors (aged between 13 and 18) as potential users, since parents' consent was required to establish a valid contractual obligation. Therefore, the Garante concluded that as of 30 March 2023, OpenAI had not implemented the necessary measures to ensure that the subscription process for ChatGPT was GDPR compliant, thereby violating Article 24 GDPR. Additionally, OpenAI failed to adopt the technical and organisational measures required to effectively uphold data protection principles, such as data minimisation, which constitutes a breach of Article 25 GDPR.

## 5. The Informational Campaign: Article 83(5)(e) GDPR

As outlined in the previous section, with Decision No. 114/2023, the Garante required OpenAI to carry out an informational campaign of a non-promotional nature across all major Italian mass media channels (radio, television, newspapers, and the Internet), to be agreed upon with the Garante, as a condition for suspending the provisional limitation on processing. The authority's request had manifold goals: (i) to inform individuals about the likely collection of their personal data for the training of algorithms; (ii) the publication of a dedicated detailed notice on the company's website; and, (iii) the availability on OpenAI's website of a tool through which all data subjects could request and obtain the deletion of their personal data.

Although OpenAI had informed the Authority that it had complied with the request, the Garante strongly opposed the campaign. The opposition was based on the fact that the campaign's terms and conditions were not agreed upon with the Authority. Additionally, the choice of media and communication methods, along with the extremely limited duration of the campaign, were deemed inadequate for effectively reaching the general public affected by ChatGPT. This failure hindered the public's ability to understand their rights, especially their right to object to the processing of personal data conducted by the company, including data used for training its models. As a result, OpenAI violated Article 83(5)(e) GDPR by not complying with an order from a Supervisory Authority pursuant to Article 58(2)(d) GDPR.

## 6. The Accuracy Principle: Article 5(1)(d) GDPR

As highlighted in the final report produced by the EDPB task force, although the purpose of the data processing is to train the GPT model rather than to provide accurate information—since the probabilistic nature of the system leads the model to generate partial or biased results – it is nonetheless likely that ChatGPT's outputs will be perceived as accurate by end users, regardless of their actual accuracy.[16] Therefore, the Garante emphasised the importance of OpenAI to provide clear information about the probabilistic mechanisms behind output generation and their limited reliability, including an explicit reference to the fact that the generated text, while syntactically correct, may be distorted or biased.

Despite the comprehensive technical and organisational measures that OpenAI has implemented at various stages of processing (such as efforts to identify and remove inaccurate or potentially harmful information during training, instructing models to refuse to provide sensitive information about individuals post-training, and allowing users to report inaccuracies and request corrections during usage), the Garante classified the breach of Article 5(1)(d) GDPR as a continuous violation and, therefore, still ongoing as of 15 February 2024, when OpenAI set up an establishment in the EU. As of that date, the one-stop-shop mechanism must be applied, resulting in the transfer of competence to the lead supervisory authority (LSA). Since the Garante could not proceed due to a lack of jurisdiction, it ordered the case files to be transferred to the Irish LSA.

## 7. Corrective and Sanctioning Measures

As outlined in the preceding paragraphs, OpenAI has engaged in multiple actions that constitute several violations of the GDPR. The legal basis, transparency, and age verification violations have been jointly considered based on the principle of the unity of action pursuant to Article 83(3) GDPR. Conversely, the violations related to the data breach and the failure

16 EDPB, 'Report of the work undertaken by the ChatGPT Taskforce' (23 May 2024), 8.

to comply with an order from the Authority constitute separate infringements. For the calculation of the administrative fine, the Garante considered three factors: (i) the nature of the violation pursuant to Article 83(4)–(6) GDPR; (ii) the severity of the violation; and, (iii) OpenAI worldwide annual turnover. At the same time, the Garante took into account the adoption of corrective measures by OpenAI to remedy the violation and mitigate its potential adverse effects as mitigating factors under Article 83(2)(f) GDPR.

In light of the above, the Garante imposed an administrative fine pursuant to Article 58(2)(i) GDPR of €15 million on OpenAI, amounting to 1.58% of the company's worldwide annual turnover of 2023. The Garante also ordered OpenAI to carry out a 6-months institutional communication campaign on radio, television, newspapers and the Internet per Article 166(7) of the Italian Privacy Code. Interestingly, this was the first instance in which the Italian DPA employed such an enforcement mechanism. The content of the campaign must be agreed with the Garante and shall promote public awareness of the functioning of ChatGPT, in particular on collecting user and non-user data for the training of generative AI and the rights exercised by data subjects, including the rights to object, rectify, and delete their data. The campaign will start within 45 days from the notification of the Garante's approval of the communication plan which OpenAI must submit within 60 days from the notification of this decision.

## IV. Comment

This decision concludes the nearly two-year dispute between OpenAI and the Italian Supervisory Authority, which began in March 2023 with a temporary ban on data processing that led to an official investigation. Arguably, among the various legal issues tackled by the Garante's decision, three stand out for their actuality in the privacy and data protection debate around LLMs in Europe, namely: i) the appropriate legal basis for the processing of personal data, in particular for training LLMs; ii) the meaningful implementation of the transparency principle; and, iii) the adoption of effective age verification systems.

OpenAI based its processing operations for service provision on the legal basis of contract performance, while it relied on legitimate interest for the processing operations related to algorithm training. However, the Garante did not assess the suitability of legitimate interest as a legal basis for processing as this matter would fall under the competence of the Irish supervisory authority as the LSA, since it concerns an ongoing violation.[17] Thus, the Italian DPA merely demonstrated that OpenAI could not prove that it had identified a legal basis before starting the processing of personal data. Among the various legal bases available to data controllers pursuant to Article 6 GDPR, legitimate interest emerges as the most functional legal basis to train a large language model. The EDPB, recalling its Guidelines on legitimate interest,[18] recently shed light on the three steps underlying the legitimate interest assessment, in particular, with regard to the context of the development and deployment of AI models. First, the interest pursued by the controller or a third party must be lawful, clear, and present; second, the processing must be necessary for the purpose of the legitimate interest pursued (this entails assessing whether the processing allows the pursuit of the purpose and whether there is no less intrusive way of pursuing this purpose); and third, a balance must be struck between the interests and fundamental rights of the data subjects and the interests of the controller.[19] For example, the EDPB considers developing LLMs to enhance cybersecurity may constitute a legitimate interest.[20] Nonetheless, the assessment that the data controller must perform—following the principle of accountability—should always take into account the specific context of processing when developing an AI model, as the outcome of the test may yield different results.

Conversely, the decision provides detailed insights on how to effectively apply the principle of transparency in the development and use of an AI model. Specifically, the question is to what extent data subjects, both users and non-users of the service, can understand the characteristics and functioning of the system, as well as the details of the processing (eg what data is collected, for what purposes, etc) so that they can exercise their rights under the GDPR. De-

17 Garante per la protezione dei dati personali (n 1), 31.

18 EDPB, 'Guidelines 1/2024 on processing of personal data based on Article 6(1)(f) GDPR', adopted on 8 October 2024.

19 EDPB, 'Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models', adopted on 17 December 2024', 21 ff.

20 Ibid, para 69.

spite OpenAI adopting several measures to uphold the transparency principle, such as improving the privacy policy, or creating a pop-up to inform users that their chats might be used to improve the service, the Garante found that the principle of transparency had been violated, given the severity of the lack of information for both users and non-users, who were unaware of the processing of their personal data. A (6-months) non-promotional informational campaign, the contents of which would need to be agreed upon with the Supervisory Authority, could serve as an effective and proportionate measure to meaningfully inform data subjects—not just the users of a specific service utilising an LLM—about the potential collection and processing of their personal data for training AI models. Given the complexities of the processing and the underlying technology, the information must be presented in an accessible and understandable way, for all data subjects.[21]

The fact that anyone can potentially use a service based on an LLM brings us to the last issue, that is, age verification mechanisms. It is important to note that the Garante did not assess the measures adopted by OpenAI through its partnership with Yoti Ltd. At the same time, the Garante acknowledged the absence of a common standard capable of ensuring, with absolute certainty, the effectiveness of a user age verification model, considering the ongoing discussion at the European level. Thus, the Garante limited itself to observe that, until 30 March 2023, these mechanisms had not been implemented, as OpenAI finalised its agreement with Yoti Ltd. on 29 September 2023. It might be beneficial to examine closer the mechanism OpenAI implemented through Yoti app. This system involves three ways OpenAI can verify users' age. First, users provide an ID document and a photo to Yoti only once, after which they can gain access to OpenAI by scanning a QR code from OpenAI's webpage; second, users take a selfie using the Yoti app or website, which estimates their age with an absolute error rate of 1.79 years; third, users scan an ID document and take a selfie: Yoti then verifies the match between the two using an automated processing system.[22] It remains to be seen whether such an age verification system will pass the assessment of European data protection authorities in the future.

## V. Conclusion

The Italian Data Protection Authority's decision to sanction OpenAI for infringing several GDPR provisions related to the development and use of the ChatGPT service goes beyond a mere enforcement action against a tech company. Against the background of upcoming data protection disputes involving EU data protection authorities and providers of AI models, this decision marks the first instance in which a European supervisory authority has addressed crucial data protection issues surrounding the debate on generative AI and LLMs, such as the accuracy of the LLMs' outputs, how to meaningfully implement the transparency principle as well as the transparency rights, the adequate and proportionate technical and organisational measures to uphold data protection principles.

21 EDPB (n 19), para 63.

22 Garante per la protezione dei dati personali (n 2), 6.