noyb – European Center for Digital Rights
Goldschlagstraße 172/4/3/2
1140 Vienna
Austria

Österreichische Datenschutzbehörde (DSB)
Barichgasse 40-42
1030 Wien

Per E-Mail: dsb@dsb.gv.at

Vienna, 29.04.2024

noyb Case   C-078

Complainant: ███████████████████████

represented under   *noyb* – European Center for Digital Rights
Article 80(1) DSGVO by: Goldschlagstraße 172/4/3/2, 1140 Vienna
                                                Austria

Respondent: **OpenAI OpCo, LLC**
                                                3180 18ᵗʰ Street, 94110 San Francisco (CA)
                                                United States

Regarding: Principle of accuracy – Article 5(1)(d) GDPR
                                                Access request – Article 15 GDPR

COMPLAINT

noyb – European Center for Digital Rights | Goldschlagstr. 172/4/3/2, 1140 Wien, Austria | ZVR: 1354838270
www.noyb.eu | general requests: info@noyb.eu   ████████████████████

page 1 of 6

1. REPRESENTATION

1. *noyb* – European Center for Digital Rights is a not-for-profit organisation active in the field of the protection of data subjects' rights and freedoms with its registered office in Goldschlagstraße 172/4/2, 1140 Vienna, Austria, registry number ZVR: 1354838270 (hereinafter: *„noyb")* (**Attachment 1**).

2. *noyb* is respresenting the complainant under Article 80(1) GDPR (**Attachment 2**).

2. FACTS PERTAINING TO THE CASE

3. The controller created and manages ChatGPT, an Artificial Intelligence (AI) application that interacts with the user by providing replies to prompts formulated as questions in any natural language of choice. ChatGPT makes use of large language models to provide answers that aim to be correct and updated. These large language models calculate the statistical probability that a given combination of words will appear in a given context and display the most likely results accordingly. This operation is possible via the use of vast datasets that the controller uses to train and "feed" the algorithm. These datasets make use of personal data, too.

4. When asking ChatGPT to provide the data subject's date of birth, the algorithm gives various inaccurate information (**Attachments 3a and 3b**). This can be easily verified by anyone with a ChatGPT account. As the data subject is a public figure, some data concerning him are online, his birthdate is not. Therefore, ChatGPT tries to infer his date of birth, but it does not provide accurate results. For example, ChatGPT stated that the data subject's day of birth was ███████████████████████████████████ As a matter of fact, the correct date would be ███████████████████████████████

5. On 04.12.2023, the data subject filed an access and erasure request with the controller. The erasure request concerned specifically the data subject's incorrect date of birth from the results displayed by ChatGPT (**Attachment 4**).

6. Concerning the access request, in its 07.02.2024 reply, the controller focused on the data subject's <u>account data</u> as a user of ChatGPT. No information has been provided about the processing of personal data performed by the relevant filing system, namely the <u>ChatGPT large language model</u>, for example with regard to what data was used to train the algorithm.

7. Concerning the erasure request, the controller stated that there is no way to prevent its systems from displaying the data subject's inaccurate date of birth in the output, if the user asks for this piece of information.

8. The controller acknowledged to have filters enabling to <u>block displaying</u> personal data of people requesting it. However, the controller stressed that it is not possible to block the data subject's date of birth without affecting other pieces of information that ChatGPT would display about him. In other words, <u>the blocking function would necessarily be all-encompassing</u> for any request on ████████████ or alike.

9. The controller does not seem to have any option to actually correct false information, but can only "hide" at the final output stage of the processing. Even if all data would be blocked, the false information would still be present in the system – just not shown to users.

10. The controller also stated that, as the data subject is a public figure, limitations in the information provided by ChatGPT concerning him would violate the controller's freedom to inform and the general public's right to be informed.

11. In other words: ChatGPT cannot correct information, cannot selectively block information and any data subject must simply life with that situation – according to the controller. ChatGPT seems to take the view that it can simply spread false information and is (other than any media company or other controller) not liable for it.

3. COMPETENT AUTHORITY

12. According to the controller's own wording, the processing consists in *"[generating] responses to user requests by predicting the next most likely words that might appear in response to each prompt"* (**Attachment 5**). This is ultimately the core of ChatGPT service. The processing is possible thanks to the use of large language models and what we assume to be a single neural network.

13. To our best understanding, this network is operated by Open AI LLC, a US-based company with seat in San Francisco, California.

14. In addition, it is worth noticing that OpenAI takes pride in preserving its nature of non-profit organisation. OpenAI explicitly declares that its for-profit branches are directly or indirectly controlled by OpenAI, Inc. 501(c)(3) Public Charity ("OpenAI Nonprofit"), a US-based entity.[1] We assume that such a control extends to the processing of personal data in the context of the large language models, as this processing is at the core of OpenAI's activities and services. Therefore, a large portion – if not all – of the decision-making concerning purposes and means of the processing takes place in the US.

15. The fact that OpenAI has recently developed an establishment in Europe (OpenAI Ireland Limited, 1st Floor, The Liffey Trust Centre, 117-126 Sheriff Street Upper, Dublin 1, D01 YC43, Irland) does not change the control over the Large Language Model. From the information available online, it results that this company is in a shared office ("Liffey Trust Center", that "*provides management consultancy for the first year*" and "*provides work space if available in Dublin with reduced rent for new businesses while they are being established*" according to its website)[2]. We are not aware of any substantial operation in Ireland, let alone any substantial decision power over the relevant processing operation (the ChatGPT Large Language Model).

16. OpenAI's Irish office opened only on 13.09.2023 – almost one year after the explosion of ChatGPT as service on the international markets – and it is probably just a *pro forma* office following the enforcement actions taken by the Italian supervisory authority at the beginning of 2023, to allow OpenAI to sneak under the "protection by inaction" of the Irish DPC.

---

[1] https://openai.com/our-structure

[2] https://www.liffeytrust.ie/

17. Consequently, the Italian supervisory authority itself did not attribute any relevance to the opening of this Irish office, as it subsequently started another investigation concerning OpenAI's latest product – "Sora" – on 08.03.2024 and did not applying the one-stop-shop.[3]

18. Even if OpenAI Ireland Limited would take some decisions on the purposed and means of the Large Language Model, it would be at most, a "joint controller" with OpenAI LLC and its non-profit controlling company, which are both based in the United States. Given that the complainant can choose to file a complaint only against one of multiple "joint controllers", such a joint controllership is irrelevant when it comes to the jurisdiction of a DSB in relation to the opponent in this complaint.

19. The GDPR allows for the filing of complaints under Article 77 against only one of the joint controllers. The data subject has decided to file his complaint only against OpenAI LLC, retaining the right to also enforce his rights against the Irish entity at any time, if it became clear that the latter has actual decision-making powers.

20. The data subject resides and works in Austria and the Österreichische Datenschutzbehörde is thus competent to handle his complaint under Articles 55 and 77 GDPR.

4. GROUNDS FOR THE COMPLAINT

4.1. Violations

21. The respondent violated the following provisions of the GDPR:

    (a) Complainant's right of access: the controller failed to provide a full reply to the data subject's access request and thus violated Articles 12(3) and 15 GDPR.

    (b) Respondent's obligation to respect the principle of accuracy: the controller does not guarantee an accurate processing of personal data concerning the data subject, thereby violating Article 5(1)(d) GDPR.

4.2. Violation of Articles 12(3) and 15 GDPR

22. The original access request was made on 04.12.2023, which means almost four months before the filing of the present complaint. So far, the respondent has provided the data subject only with general information concerning data processed in the context of the complainant's <u>user account</u>. There is no response when it comes to personal data processed through the <u>ChatGPT large language model</u>. Overall, the controller has simply not replied in relation to the relevant filing system.

23. It is manifest from the results displayed by ChatGPT to any user prompting the software with questions concerning the complainant that the controller processes much more personal data about the complainant. If the large language model did not store and process a considerable amount of personal data regarding the complainant, it could not provide the users with meaningful answers on him. The controller itself acknowledges that the data subject is a public

---

[3] https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9991867#english

figure and informing the public about his professional history and activity is part of the controller's freedom to inform.

24. However, the processing of data in the context of the large language models that enable ChatGPT to function has not been addressed at all by respondent. The data subject has not received any information on what data concerning him is processed by OpenAI, for the purpose of powering ChatGPT, where this data came from, what the legal basis may be, how long ChatGPT plans to store this information or who it provided this (false) information to.

25. Therefore, in not providing the required information within the legally established deadline, OpenAI violated Articles 12(3) and 15(1) to (3) GDPR.

4.3.   Violation of Article 5(1)(d) GDPR

26. Article 5(1)(d) GDPR imposes on the controller an obligation to erase or rectify without delay inaccurate data.

27. The controller was made aware of the accuracy issue by the data subject. However, no measure has been taken in this regard. When asked to provide the complainant's data of birth, ChatGPT keeps displaying inaccurate information.

28. The respondent states that the only way to prevent the inaccurate information from appearing would be to block *any* information concerning the data subject. This would in turn violate the controller's freedom to inform and the general public's right to be informed, as the data subject is a public figure. Thus, the controller refused to take action.

29. In the present case, however, the respondent cannot invoke its freedom of expression. In the area of data protection, untrue statements of fact do not fall within the scope of protection of freedom of expression. Furthermore, the (false) date of birth of the complainant would not contribute anything to a debate of public interest. The respondent was also unable to cite any legal provision under Article 85 GDPR that would allow a departure from the principle of accuracy in favour of the complainant's right to express an opinion - which is not applicable here.

30. It must be stressed that the claimed technically impossibility to erase or rectify the data subject's date of birth without blocking other relevant pieces of information is by no means a valid justification to derogate to the principle of accuracy under Article 5(1)(d) GDPR. The fact that a software developed by a controller is unable to comply with the law makes the processing simply unlawful – but never the law inapplicable.

31. Therefore, as long as ChatGPT keeps showing inaccurate data on the respondent, the controller violates Article 5(1)(d) GDPR.

5. REQUESTS AND SUGGESTIONS

5.1. Request to investigate

32. The complainant invites the competent authority to use its powers under Article 58(1) GDPR to investigate the processing and in particular the measures taken by the defendant to guarantee the accuracy of personal data processed in the context of its large language models.

5.2. Request for a declaratory decision

33. The complainant requests that the complaint be upheld and that OpenAI be found to have infringed Article 5(1)(d) and Article 15 GDPR.

5.3. Request to take appropriate corrective measures

34. The complainant requests the competent supervisory authority to make use of its powers under Article 58(2)(c) and (d) and order the respondent to:

(a) comply with the complainant's access request pursuant to Article 15 GDPR;

(b) bring processing operations into compliance with the principle of accuracy under Article 5(1)(d) GDPR, in particular by erasing or rectifying the complainant's date of birth.

5.4. Suggestion to impose a fine

35. Finally, the complainant invites the competent supervisory authority – on the basis of its powers under Articles 58(2)(i) and 83 GDPR – to impose an effective, proportionate and dissuasive administrative fine in the amount necessary to guarantee the controller's future compliance with the GDPR.

6. CONTACT

36. Communications between *noyb* and the DSB in the course of this procedure can be done by email at ███████████ with reference to the **Case C-078** or by phone at ███████████.