GP Partners
Gawronski, Biernatowski Sp.K. info@gppartners.pl
al. Jana Pawła II 12
00-124 Warsaw

Warsaw, August 29, 2023.

Complainant:
Lukasz Olejnik

represented by:
r. pr. Maciej Gawronski
GP Partners Gawronski, Biernatowski Sp.K.
al. Jana Pawła II 12, 00-124 Warsaw

President of the Personal Data Protection Office
Stawki 2 Street
00-193 Warsaw

COMPLAINT
AGAINST UNLAWFUL PROCESSING OF PERSONAL DATA

Acting on behalf of the complainant, Mr. Łukasz Olejnik (power of attorney with proof of payment enclosed) I hereby

complain

on the unlawful[1] processing of **Mr. Łukasz Olejnik's** personal data by **OpenAI OpCo, LLC**, 3180 18th Street, San Francisco, CA, USA, within the ChatGPT tool,

involving processing in violation of the principles of lawfulness, fairness, and transparency, i.e., in violation of Article 5(1)(a) GDPR, failure to exercise the right of access to personal data and information about the processing of personal data pursuant to Article 15 in conjunction with Article 12 GDPR, failure to exercise the right to rectify personal data pursuant to Article 16 in conjunction with Article 12 GDPR, and processing in a manner contrary to the principle of data protection by design, i.e., in violation of Article 25(1) GDPR

and apply:

1) to initiate administrative proceedings regarding the processing of Mr. Łukasz Olejnik's personal data unlawfully by OpenAI OpCo, LLC,
2) to require OpenAI OpCo, LLC to exercise, in accordance with Articles 15 and 16 in conjunction with Article 12 GDPR, the rights of Mr. Łukasz Olejnik, i.e. the right of access to personal data and to information about the processing of personal data and the right to rectify personal data,
3) to require OpenAI to submit a data protection impact assessment (DPIA) document to the President of the Personal Data Protection Office regarding the processing of personal data for purposes related to the provision of the ChatGPT tool.

In the present case, the *one-stop-shop* principle under Article 56 in conjunction with Article 60 GDPR does not apply (this follows from the logic presented by the French supervisory authority,

[1] I.e., not in compliance with the provisions of 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 (General Data Protection Regulation), ("**GDPR**").

i.e., CNIL, in its January 21, 2019 decision to impose a penalty on Google[2]) - more on this in Section 2.2 of the complaint.

REASONING

1. **FACTS**

1.1. OpenAI and ChatGPT

OpenAI. OpenAI OpCo, LLC ("**OpenAI**") is an organization that conducts research on the development of artificial intelligence and the implementation of artificial intelligence-based tools. OpenAI was established in 2015 as a non-profit research center to promote human-friendly artificial intelligence. As of 2019, there is also a company, OpenAI LP, which develops artificial intelligence tools for commercial purposes.

ChatGPT. Among the most well-known and popular (in terms of number of users) tools is **Chat Generative Pre-Trained Transformer**, or ChatGPT for short. ChatGPT is a chatbot application used to generate responses to user input into the chat in the form of prompts, messages. ChatGPT is a multi-purpose tool, mainly used for generating content, holding conversations on a variety of topics and answering users' questions.

ChatGPT currently uses the Generative Pre-trained Transformer 3,5 and 4 (GPT-3,5 and GPT-4) language model. The GPT-3.5 and GPT-4 models were developed using a machine learning technique, using *big data sets*. The models allow ChatGPT to compile a large amount of commonly available data to generate content relevant to prompts, user queries. The basic scheme of ChatGPT's operation is as follows:

> *It* [Chat-GPT] *generates texts using a computational technique called a "transformer-type neural network," and the network's parameters were determined by "training" it in advance on examples from a huge text database. Generally, as input, the program receives a fragment of natural language text, such as a query, and is tasked with generating another text, sensible and grammatically correct, that best fits as a continuation of the given fragment. The match is determined by the text database on which the program was trained. It can be said that the program generates the most likely continuation of the text based on the texts present in the database[3].*

The results generated by ChatGPT (including their correctness and validity) are highly dependent on the data ChatGPT has been fed, as well as the effectiveness of the language models. The databases that ChatGPT uses to generate texts come from a wide variety of sources and cover a variety of data categories, including personal data.

1.2. Processing of Mr. Łukasz Olejnik's data by OpenAI

The complainant in the present case, Mr. Łukasz Olejnik, is an independent researcher dealing with, among other things, cyber security, privacy and data protection. He is also the author of works in these areas.

Mr. Łukasz Olejnik addressed a question to ChatGPT in order to generate his own biography. In response, ChatGPT generated a text with a short biography and description of Mr. Łukasz Olejnik's career. The text generated by ChatGPT contained partially false information. Among other things, ChatGPT indicated:

> *Łukasz Olejnik has done extensive research on web browsing history and related topics such as online tracking, user privacy, and web security. He has written numerous articles and academic papers on the subject, and has been recognized for his contributions to the field. For example, in 2020, he co-authored a paper titled "What Web Browser History Tells Us About User Activity and Privacy."*

In the above excerpt, ChatGPT indicated the wrong title and date of an article authored by Mr. Łukasz Olejnik. In another generated excerpt, ChatGPT incorrectly attributed the authorship of several articles to Mr. Łukasz Olejnik.

---

[2] Délibération de la formation restreinte n° SAN - 2019-001 du 21 janvier 2019 prononçant une sanction pécuniaire à l'encontre de la société X, https://www.legifrance.gouv.fr/cnil/id/CNILTEXT000038032552/

[3] A. Kisielewicz, Artificial Intelligence Fairy Tales and Real Dangers, https://wszystkoconajwazniejsze.pl/andrzej-kisielewicz-bajki-o-sztucznej-inteligencji-i-prawdziwe-zagrozenia/ (accessed 03/08/2023).

Exhibit 1: Screenshot 1 containing an excerpt from a conversation between Mr. Łukasz Olejnik and ChatGPT

Mr. Łukasz Olejnik, in addition to questions about his biography, asked ChatGPT an additional question about his gender. ChatGPT responded that Mr. Łukasz Olejnik is male. In order to understand how this information was determined and generated, Mr. Łukasz Olejnik asked how ChatGPT determined his gender. In response, ChatGPT described the process that led to the generation of the answer. The description shows, among other things, that ChatGPT analysed available information, including photos, biographical information, including gender from several different sources.

Exhibit 2: Screenshot #2 containing an excerpt from a conversation between Mr. Łukasz Olejnik and ChatGPT

1.3. Request of Mr. Łukasz Olejnik to exercise data subject's rights and exchange of correspondence with OpenAI[4]

Request to exercise individual's rights. In view of the identified errors in the information generated by ChatGPT, Mr. Łukasz Olejnik sent an email to OpenAI on March 27, 2023 with a request:

1) to indicate any information concerning Mr. Łukasz Olejnik processed by OpenAI in accordance with Articles 12, 14 and 15 GDPR,
2) to correct information regarding the article, which is titled "What Web Browser History Tells Us About User Activity and Privacy" generated by ChatGPT.

In his correspondence with OpenAI, Mr. Łukasz Olejnik explicitly mentioned the provisions of Articles 12, 14 and 15 GDPR as the legal basis for the request mentioned in point 1 above. In addition, Mr. Łukasz Olejnik mentioned Articles 13.2.f), 14.2.g) and 15.1.h) GDPR as an example of the information he specifically wanted from OpenAI in response to his request.

> *I would kindly **ask about all the information you have concerning Łukasz Olejnik** (the user in your system identified via email lukasz.w3c@gmail.com, but also the external data that you used). Concretely, since I am a citizen of the EEA, specifically the EU, please consider the principles as drawn directly from the GDPR. **With a specific attention to article 12, article 14, article 15.***
>
> *I would also ask you **to change the following bio**: "Łukasz Olejnik has done extensive research on web browsing history and related topics such as online tracking, user privacy, and web security. He has written numerous articles and academic papers on the subject, and has been recognized for his contributions to the field. For example, in 2020, "he co-authored a paper titled "What Web Browser History Tells Us A "out User Activity and Privacy.""*
>
> *The mentioned title of the paper is incorrect, as well as the date.*
>
> *(...)*
>
> *I expect you to provide me the appropriate "logic involved in any automatic personal data processing", and the provisions of article 13(2)(f), not to mention 14(2)(g), and 15(1)(h).*

Exhibit 3: Email correspondence between Mr. Łukasz Olejnik and OpenAI regarding the processing of personal data within ChatGPT.

OpenAI's response. OpenAI has not complied with the requests made by Mr. Łukasz Olejnik in correspondence dated March 27, 2023. In the first message, OpenAI stated that it considers the request made by Mr. Łukasz Olejnik to have been executed by generally blocking ChatGPT's ability to respond to questions containing his name. In a subsequent response in an email message dated April 18, 2023. OpenAI indicated that it is not possible to correct the information about Mr. Łukasz Olejnik in the fragment of text generated by ChatGPT.

---

[4] In order to present the actual exchange of correspondence, the complaint cites excerpts from the messages in English. Attached to the complaint we provide a transcript of the original correspondence (in English) and, in addition, a machine translation of the entire correspondence.

Below is an excerpt from correspondence from OpenAI:

> (...) *As some background, ChatGPT is designed to produce conversational text by predicting and outputting the next most likely word in response to a user's request. In some cases the next most likely word may also not be the most accurate one. We are working to improve the accuracy of our models, but we warn users that ChatGPT output may be false or factually inaccurate in the ChatGPT UI as well as Section 3(d) of our Terms of Use. Although* ***we are unable to change the information in the statement you have flagged****, we can address the issue by preventing your name from being generated by ChatGPT. Please let us know if you would like us to do this.*

In addition to the above information, OpenAI also pasted a link to its Privacy Policy into the correspondence and attached a PDF document titled "*OPENAI DATA SUBJECT ACCESS REQUEST RESPONSE*." The document contained information about OpenAI's data processing in response to the data access request.

Exhibit 3: Email correspondence between Mr. Łukasz Olejnik and OpenAI regarding the processing of personal data within ChatGPT.

Further exchange of correspondence. In response to the above information received from OpenAI dated April 19, 2023. Mr. Łukasz Olejnik indicated what additional information about OpenAI's processing of his data he would like to receive (including the categories of data processed and the sources from which OpenAI obtained the data and the categories of data recipients).

In correspondence dated May 6, 2023. OpenAI indicated that it processed Mr. Łukasz Olejnik's personal data by automated means. Although OpenAI's response was partially misleading, according to the correspondence, "*OpenAI does process any personal data using automated decisions with a legal or similarly significant effect*." Later in the message, OpenAI indicates:

> *This is due to the fact that ChatGPT is designed to provide a response to a prompt provided by a human user, and therefore there is always human involvement. As a result, Article 22 of the GDPR is not applicable to OpenAI's processing activities, so there is no requirement to notify data subjects of the information described in Articles 13(2)(f), 14(2)(g) and 15(1)(h) of the GDPR.*

It would appear from the latter passage that there is no data processing involving automated decision-making by OpenAI. The information provided by OpenAI was therefore contradictory and could lead to different, opposite conclusions. The above information was subsequently amended by OpenAI to indicate that OpenAI does not make automated decisions when processing personal data.

In the same email, OpenAI provided additional information on the entities sub-processing data on OpenAI's behalf (sub-processors) and additional security mechanisms for the processed data (additional data controls), by pasting links to websites in OpenAI's web domain.

In a subsequent OpenAI email dated June 7, 2023, in response to additional questions from Mr. Łukasz Olejnik, OpenAI provided further general information on how the GPT-3.5 and GPT-4 models work, personal data sources and OpenAI affiliates. The message dated June 7, 2023 was the last of the messages Mr. Łukasz Olejnik received from OpenAI in response to his original request of March 27, 2023.

Exhibit 3: Email correspondence between Mr. Łukasz Olejnik and OpenAI regarding the processing of personal data within ChatGPT.

1.4. Doubts about the level of protection of personal data processed by OpenAI

A relevant factual context of the present case regarding the processing of Mr. Łukasz Olejnik's personal data by OpenAI are the doubts that have arisen about the level of protection of personal data processed within the framework of artificial intelligence tools, in particular ChatGPT.

A manifestation of these concerns are the efforts that the European supervisory authorities[5] , have made to understand and address the risks that ChatGPT and other similar tools pose to

[5] ChatGPT is entering a world of regulatory pain in Europe, https://www.politico.eu/article/chatgpt-world-regulatory-pain-eu-privacy-data-protection-gdpr/ (accessed 04.08.2023).

individuals' personal data and privacy. Concerns about the security of personal data processed by ChatGPT are triggered in particular by the exponential growth of interest in the tool and its broad applicability, as well as the massive amount of data collected to train models within ChatGPT by OpenAI.

Below I indicate examples of the actions and positions of the supervisory authorities:

- Italian supervisory body Garante per la Protezione dei Dati Personali, which initially blocked general access to ChatGPT, indicated, among other things, that OpenAI has no legal basis to collect and store personal data en masse for the purpose of training ChatGPT, the data processed is not correct, and OpenAI does not check the age of ChatGPT users[6] . ChatGPT was later unblocked due to some improvements made by OpenAI on privacy and personal data processing[7].
- French regulator CNIL has announced that it will intensify its data protection activities in connection with *generative AI* tools in 2023. In its four-point AI action plan, it pointed out that the most important challenge in designing and using tools such as ChatGPT is to ensure the protection of personal data[8].
- The Hessian Data Protection Commissioner in Germany, in turn, addressed a series of questions directly to OpenAI, pointing out the unclear purpose of ChatGPT's data processing and the unclear sources from which ChatGPT draws its knowledge[9].
- The ICO's UK regulator, in its general communication dated April 3, 2023, noted that:

  Organisations developing or using generative AI should be considering their data protection obligations from the outset, taking a data protection by design and by default approach. This isn't optional – if you're processing personal data, it's the law[10].

- In addition, G7 countries have developed a common position on artificial intelligence. The statement identifies a number of areas where data protection regulators believe generative artificial intelligence tools may pose risks. These areas include the legal basis for processing personal data, security of personal data, transparency of processing, accountability of controllers and minimization of processing[11].

The cited positions of supervisory authorities on the lawfulness of processing of personal data within the framework of generative artificial intelligence tools confirm the doubts that arise in connection with the processing of Mr. Łukasz Olejnik's personal data by OpenAI in the present case.

2. LEGAL REASONING

2.1. Application of GDPR to the processing of personal data of Mr. Łukasz Olejnik by OpenAI

---

6 Original decision dated March 30, 2023 by the Garante per la protezione dei dati personali (https://garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9870832, accessed August 01, 2023).

7 ChatGPT: OpenAI riapre la piattaforma in Italia garantendo più trasparenza e più diritti a utenti e non utenti europei, https://garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9881490 (accessed August 01, 2023).

8 Artificial intelligence: the action plan of the CNIL, 16 May 2023, https://www.cnil.fr/en/artificial-intelligence-action-plan-cnil, (accessed August 01, 2023).

9 Pressemitteilung, Anhörung, Hessischer Datenschutzbeauftragter fordert Antworten zu ChatGPT https://datenschutz.hessen.de/presse/hessischer-datenschutzbeauftragter-fordert-antworten-zu-chatgpt, (accessed 01.08.2023).

10 Generative AI: eight questions that developers and users need to ask, own translation, https://ico.org.uk/about-the-ico/media-centre/news-and-blogs/2023/04/generative-ai-eight-questions-that-developers-and-users-need-to-ask/, (accessed August 01, 2023).

11Roundtable of G7 Data Protection and Privacy Authorities Statement on Generative AI, June 21, 2023, https://www.ppc.go.jp/files/pdf/G7roundtable_202306_statement.pdf, (accessed August 01, 2023).

The provisions of GDPR apply to the processing of Mr. Łukasz Olejnik's personal data by OpenAI to the extent described in point 1 of the complaint.

Material scope of application. According to Article 2(1) GDPR:

> *This Regulation applies to the processing of personal data wholly or partly by automated means and to the processing other than by automated means of personal data which form part of a filing system or are intended to form part of a filing system.*

As can be seen from the description of the facts in point 1 - OpenAI processed and processes the personal data of Mr. Łukasz Olejnik.

At the same time, none of the exemptions described in Article 2(2) GDPR apply to the processing of Mr. Łukasz Olejnik's personal data by OpenAI.

Territorial scope of application. According to Article 3(2) GDPR:

> *This Regulation applies to the processing of personal data in the context of the activities of an establishment of a controller or a processor in the Union, regardless of whether the processing takes place in the Union or not.*

As can be seen from the Privacy Policy provided on the OpenAI website[12], OpenAI also has a business unit based in Ireland. This indicates that the processing of Mr. Łukasz Olejnik's personal data occurred in connection with the activities carried out by OpenAI's business unit in the Union, although the controller of the personal data deciding the processing is OpenAI, based in the US.

The processing of Mr. Łukasz Olejnik's personal data is therefore covered by the territorial scope of the GDPR.

2.2. Exclusion of the application of the one-stop-shop principle

Despite the fact that OpenAI has a representative, an organizational unit in the EU - the *one-stop-shop* mechanism described in Article 56 GDPR, in conjunction with Article 60 GDPR, does not apply to OpenAI's cross-border data processing.

In the present case, the logic presented by the French supervisory authority, i.e. CNIL, in its January 21, 2019 decision to impose a penalty on Google[13] applies analogously. The French authority, in issuing the Google decision, pointed out that although Google is headquartered in Ireland, the Irish-based entity in fact "did not have decision-making powers" with regard to the purposes and means of cross-border data processing. For this reason, the CNIL decided that the *one-stop-shop* mechanism did not apply, and therefore CNIL had jurisdiction to make the decision[14].

The above logic corresponds to the circumstances of the present case - in fact, it is OpenAI, based in the US, that decides on the means and purposes of processing, in particular with regard to the processing of data within the Chat-GPT tool, which was developed by the US entity. For this reason, the *one-stop-shop* mechanism will not apply in the case, and the PUODO will have jurisdiction to decide on the merits of the present complaint.

---

[12] OpenAI Privacy policy, https://openai.com/policies/privacy-policy (accessed August 04, 2023).

[13] Délibération de la formation restreinte n° SAN - 2019-001 du 21 janvier 2019 prononçant une sanction pécuniaire à l'encontre de la société X, https://www.legifrance.gouv.fr/cnil/id/CNILTEXT000038032552/

[14] The CNIL's argument also seems to be indirectly supported by the practice of the European Data Protection Board, which, in its updated 9/2022 Guidelines on notification of personal data protection breaches, indicated that regardless of the fact that a controller outside the EU has a representative in the EU (as is the case with OpenAI), it will be that controller's responsibility to report the breach to each of the supervisory authorities in the country where the affected individuals reside. There was no such provision in the previous guidelines which indicates a partial change of direction in the understanding of the one-stop-shop mechanism (https://cowprawiepiszczy.com/2022/11/naruszenie-ochrony-danych-osobowych-najwazniejsze-nowe-wskazowki-uodo-i-erod/, accessed 11.08.2023).
Link to EROD guidelines:
https://edpb.europa.eu/system/files/2023-04/edpb_guidelines_202209_personal_data_breach_notification_v2.0_en.pdf
Original text of the guidelines: *However, the mere presence of a representative in a Member State does not trigger the one-stop shop system. For this reason, the breach will need to be notified to every supervisory authority for which affected data subjects reside in their Member State. This (These) notification(s) shall be the responsibility of the controller.*

2.3. Violations in the processing of personal data of Mr. Łukasz Olejnik

Mr. Łukasz Olejnik identifies violations in the processing of his personal data by OpenAI in the following areas:

1) violation of the fundamental principle of data processing in Article 5(1)(a) GDPR, i.e. the principle of processing data lawfully, fairly, and transparently,
2) improper execution of the data subject's rights, including the right of access to data and the right to rectification,
3) failure to ensure a sufficient level of security of processed personal data and violation of the principle of data protection by design (privacy by design).

2.4. Violation of the principle of compliance with the law, fairly and transparently (Article 5(1)(a) of the GDPR).

The primary aspect of violations in the processing of Mr. Łukasz Olejnik's personal data by OpenAI is a violation of the principle of lawfulness, fairness and transparency of personal data processing. This principle is fundamental to all obligations under the GDPR.

Pursuant to Article 5(1)(a) of the GDPR:

> *1. personal data shall be:*
>
> *(a) processed lawfully, fairly and in a transparent manner in relation to the data subject ('lawfulness, fairness and transparency');*

In practice, the principle of fairness means that data should be processed in a loyal and honest manner with respect to the data subject[15]. From the facts of the case, it appears that OpenAI systemically ignores the provisions of the GDPR regarding the processing of data for the purposes of training models within Chat-GPT, a result of which, among other things, was that Mr. Łukasz Olejnik was not properly informed about the processing of his personal data. Violations related to the provision of information on the processing of Mr. Łukasz Olejnik are described later in the compliant.

The facts of the case show that:

1) OpenAI **violated the principle of lawfulness (legitimacy) of processing**, because it is apparent from the way OpenAI approached the processing of Mr. Łukasz Olejnik's data, as well as the performance of its obligations related to the processing of his data, that OpenAI in fact ignored the provisions of GDPR with respect to processing for the purposes of (i) training models within Chat-GPT, as well as (ii) generating content by Chat-GPT (based on user questions).

   What is more, OpenAI responded to Mr. Łukasz Olejnik's requests in a **facade-like** manner, providing a series of information that lacked concrete content and explanations regarding the processing of Mr. Łukasz Olejnik's data, including within the framework of model training, to the extent required by the provisions of GDPR.

2) OpenAI **has violated the principle of fairness of data processing,** as Mr. Łukasz Olejnik's data has been and is being processed by OpenAI in an untrustworthy, dishonest, and perhaps unconscientious manner, since OpenAI is not able to comprehensively inform on this processing.

   In this context, it would be expedient for the authority, given the above violations, most likely of a systemic nature, to require OpenAI to submit a Data Protection Impact Assessment (DPIA) document. This document could be an important element in assessing whether OpenAI's processing of data within the Chat-GPT tool is in compliance with the GDPR. The DPIA prepared by OpenAI should include a version history so that it is possible to verify what changes were made and how they affected the DPIA.

3) OpenAI **violated the principle of transparency in processing** - this is evidenced primarily by the fact that Mr. Łukasz Olejnik, as a data subject, was unable to obtain comprehensive and reliable information about how his personal data was being processed, contrary to his rights. Moreover, the evidence shows that initially OpenAI informed that it does not process

[15] M. Gawronski (ed.) Personal Data Protection. Guide to the Act and RODO with templates, Warsaw 2018, p. 94.

Mr. Łukasz Olejnik's personal data by automated means, and later OpenAI provided the opposite information.

The aforementioned violations relate specifically to data processed to train the models used by ChatGPT. According to publicly available information, ChatGPT relies on source data (from various sources), but this data "ends" in 2021[16]. This must mean that OpenAI, in order to use this data later - has made a copy of it, which means that processing has occurred.

2.5. Failure to properly exercise the rights of Mr. Łukasz Olejnik as a data subject

Request to exercise rights. In an email correspondence to OpenAI dated March 27, 2023 Mr. Łukasz Olejnik requested OpenAI to exercise the following rights:

1) The right of access to personal data (Article 15 GDPR),
2) The right to rectify personal data (Article 16 GDPR).

2.5.A. Right of access to personal data

Right of access to personal data. In accordance with Article 15(1) GDPR:

The data subject shall have the right to obtain from the controller confirmation as to whether or not personal data concerning him or her are being processed, and, where that is the case, access to the personal data and the following information:

a) the purposes of the processing;

b) the categories of personal data concerned;

c) the recipients or categories of recipient to whom the personal data have been or will be disclosed, in particular recipients in third countries or international organisations;

d) where possible, the envisaged period for which the personal data will be stored, or, if not possible, the criteria used to determine that period;

e) the existence of the right to request from the controller rectification or erasure of personal data or restriction of processing of personal data concerning the data subject or to object to such processing;

f) the right to lodge a complaint with a supervisory authority;

g) where the personal data are not collected from the data subject, any available information as to their source;

h) the existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject.

The right of access regulated by Article 15 GDPR covers three aspects - the right to obtain confirmation that personal data is being processed, the right to obtain certain information about the processing of personal data, and the right to obtain access to the personal data being processed.

[16] https://rodoradar.pl/chat-gpt-i-sztuczna-inteligencja-a-przepisy-o-ochronie-danych-osobowych/ ("*Currently, ChatGPT is a closed language model, with the database (in the free version) current as of September 2021. This means that the model has been trained on the basis of information available on the web, and not only, exclusively in this particular period*.").

How OpenAI exercised the right? OpenAI, for the purpose of exercising the right of Mr. Łukasz Olejnik, provided a document entitled "OPENAI DATA SUBJECT ACCESS REQUEST RESPONSE" attached to an email correspondence dated April 18, 2023. In the document, OpenAI provided information including:

a) purposes of data processing,

b) categories of data processed,

c) categories of data recipients,

d) data retention period,

e) information about the source of the data, if the personal data was not collected from the data subject,

f) information about automated decisions,

g) data transfer information,

h) information about the rights of the data subject.

In addition, in a separate section of the document, OpenAI provided information on how to obtain copies of personal data, indicating in this regard the following categories of data to which Mr. Łukasz Olejnik can access:

- User account name and information,
- billing information,
- IP address,
- Chat history, user information and login data,
- Data *uploaded* by users (known as *uploads*),
- Communication with the support department,
- activity in the forums.

Violations. Although the described scope of information that OpenAI provided to Mr. Łukasz Olejnik gives the impression of corresponding to the requirements of Article 15 GDPR, in fact, Mr. Łukasz Olejnik's request was not executed properly, resulting in a violation of Article 15 GDPR.

Lack of information about the processing and copies of data processed to train the models. First of all, OpenAI omits from the information provided, including in the Privacy Policy available online at[17], information about the processing of personal data used to train the language models used by ChatGPT. Although OpenAI indicates that the data used to train the models includes personal data, **OpenAI does not** actually **provide any information about the processing operations involving this data**. OpenAI thus violates a fundamental element of the right under Article 15 GDPR, i.e., the obligation to confirm that personal data is being processed.

Notably, OpenAI did not include the processing of personal data in connection with model training in the information on categories of personal data or categories of data recipients. Providing a copy of the data also did not include personal data processed for training language models. As it seems, the fact of processing personal data for model training OpenAI hides or at least camouflages intentionally. This is also apparent from OpenAI's Privacy Policy, which omits in the substantive part the processes involved in processing personal data for training language models.

OpenAI reports that it does not use so-called "training" data to identify individuals or remember their information, and is working to reduce the amount of personal data processed in the "training" dataset. Although these mechanisms positively affect the level of protection of personal data and comply with the principle of minimization (Article 5(1)(c) of the GDPR), their application does not change the fact that **"training"** data **are processed and include personal data**. The provisions of GDPR apply to the processing operations of such data, including the obligation to grant the data subject access to the data and provide the information indicated in Article 15(1) of GDPR.

Insufficient information about recipients of data. The information provided to Mr. Łukasz Olejnik by OpenAI about the recipients of his personal data indicated only general categories of recipients (including vendors and service providers, OpenAI affiliates, third parties). This way of

[17] OpenAI Privacy policy, https://openai.com/policies/privacy-policy (accessed August 04, 2023).

informing about recipients of data does not meet the requirements of Article 15(1)(c) GDPR, from which it follows that the controller should provide information on:

> *the recipients or categories of recipient to whom the personal data have been or will be disclosed*

While it is not clear from the provision itself whether the controller should identify specific recipients with their identity (if possible) - such an understanding of Article 15(1)(c) GDPR is commonly accepted, and was recently confirmed in the ruling of the Court of Justice of the European Union of January 12, 2023 in Case C-154/21[18]. In response to a preliminary question, the Court indicated that:

> *Article 15(1)(c) of 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 (General Data Protection Regulation),*
>
> *must be interpreted as meaning that the data subject's right of access to the personal data concerning him or her, provided for by that provision, entails, where those data have been or will be disclosed to recipients, an obligation on the part of the controller to provide the data subject with the actual identity of those recipients, unless it is impossible to identify those recipients or the controller demonstrates that the data subject's requests for access are manifestly unfounded or excessive within the meaning of Article 12(5) of Regulation 2016/679, in which cases the controller may indicate to the data subject only the categories of recipient in question.*

It follows from the above ruling that OpenAI should have indicated in its response to Mr. Łukasz Olejnik's request **the identity of all recipients of data that OpenAI can identify**. Therefore, OpenAI's indication of only a category of selected recipients, e.g. in the case of service providers for OpenAI or OpenAI affiliates, must have violated the obligations imposed on OpenAI as a personal data controller. OpenAI has the ability to identify these recipients, so the information provided to Mr. Łukasz Olejnik was too general.

Moreover, OpenAI has not identified among the categories of recipients of personal data the **individuals using ChatGPT**. Persons directing an inquiry to ChatGPT about Mr. Łukasz Olejnik may receive his personal data from OpenAI (these persons would therefore be recipients of the data, which would have to be notified under the right of access). In this case, moreover, it would be appropriate to indicate only the categories of data recipients, as it would not be possible to identify specific recipients.

The above violations are all the more evident considering that Mr. Łukasz Olejnik explicitly indicated in his correspondence of April 19, 2023 to OpenAI that he wanted specific information about the recipients of his personal data. There could therefore be no doubt about the scope of the information requested by Mr. Łukasz Olejnik. On the other hand, OpenAI, in response to additional questions about the recipients of the data, pointed to a list of entities sub-processing personal data (so-called *sub-processors*) on behalf of OpenAI. Undoubtedly, information about sub-processors (Article 28(2) GDPR) is different information than information about recipients of personal data, and it is the latter that is covered by the right of access under Article 15 GDPR. The information provided by OpenAI was inconsistent with the content of the request in this regard, and in addition misleading.

Lack of sufficient information on data sources. It is also questionable how OpenAI has exercised its right of access to provide information on the sources of personal data, in cases where the personal data was not collected from the data subject. According to Article 15(1)(g) GDPR:

> *(1) The data subject shall have the right to obtain from the controller confirmation as to whether or not personal data concerning him or her are being processed, and, where that is the case, access to the personal data and the following information:*
>
> (...)
>
> *(g) where the personal data are not collected from the data subject, any* ***available information about as to their source****;*

[18] Judgment of the Court of 12.01.2023, C-154/21, RW v. ÖSTERREICHISCHE POST AG., LEX No. 3454592.

It is clear from the above that the controller should provide all available information about the source of personal data collected not from the data subject. In this regard, OpenAI indicated:

> *To develop and improve our services, we obtain and analyze data sets from publicly available sources, such as the Internet, or from third parties that we have previously verified.*

The quoted excerpt is the entirety of the information that OpenAI provided to Mr. Łukasz Olejnik in exercise of his right of access to data to the extent referred to in Article 15(1)(g) GDPR. It is difficult to consider that the scope of information so provided covers *all available information* about the source of personal data. In particular, it would be necessary to indicate, at the very least, which categories of sources OpenAI uses to collect the data needed to train the models, as well as from which third parties the data comes from. The way Article 15(1)(g) GDPR is worded provides a reasonable basis to expect the controller to provide at least the above information about the sources of personal data.

Conclusions. The circumstances presented prove that OpenAI violated Article 15 GDPR due to improper execution of the right of access to data. The literature[19] indicates that:

> *The absence of one of the mandatory elements of any communication with the data subject will be considered a violation of data protection laws*.

OpenAI's violation of Article 15 GDPR also violated the fundamental principle of lawfulness, fairness and transparency of processing in **Article 5(1)(a) GDPR**.

2.5.B. Right to rectify personal data

Right to rectify personal data. In accordance with Article 16 GDPR:

> *The data subject shall have the right to obtain from the controller without undue delay the rectification of inaccurate personal data concerning him or her. Taking into account the purposes of the processing, the data subject shall have the right to have incomplete personal data completed, including by means of providing a supplementary statement.*

Closely related to the right to rectification is the principle of data accuracy expressed in Article 5(1)(d) GDPR. The literature[20] indicates that:

> *Data accuracy requires us to implement procedures to verify the quality of data and take care of the quality of personal data, as well as to enable the exercise of the right to rectify and update data (Article 16 of the GDPR).*

How OpenAI exercised the right? In response to the request for rectification of personal data submitted by Mr. Łukasz Olejnik, OpenAI has indicated that it is not possible to rectify incorrect personal data covered by the request.

Instead of executing a request for rectification of personal data, OpenAI, despite the absence of such a request, restricted the processing of personal data (Article 18 GDPR) by blocking the ability to generate responses to questions directed to ChatGPT regarding Mr. Łukasz Olejnik.

OpenAI's action did not constitute deletion of data (exercise of the right to be forgotten), because Mr. Łukasz Olejnik's personal data was not deleted by OpenAI, but there was only a limitation of processing by ceasing the processing operation of making Mr. Łukasz Olejnik's personal data available to ChatGPT users. It is doubtful, moreover, whether it is at all possible for OpenAI to exercise its right to delete data processed for the purpose of training GPT models.

Violations. Accuracy of personal data is a fundamental principle of processing and an obligation of every data controller. In accordance with Article 5(1)(d) GDPR:

> *Personal data shall be:*
>
> (...)
>
> *accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay ('accuracy')*

---

[19] M. Gawronski (ed.) Personal Data Protection. Guide to the Act and RODO with templates, Warsaw 2018, p. 206.

[20] Ibid, pp. 98-99.

It follows from the above provision that if the personal data is not accurate in light of the purposes of its processing - the controller should take **all reasonable measures** to erase or rectify the data. The literature[21] indicates that:

> *Attention to the quality of processed data is intended to protect data subjects. The processing of outdated, erroneous or otherwise incorrect data may entail negative consequences for data subjects, as well as for those who process the data, which is why the EU legislator considers it essential to ensure that the processed data is correct, i.e. factually correct, up-to-date and free of errors.*

Accuracy of data is a principle, but according to Article 5(1)(d) GDPR, this principle is not absolute. The inaccuracy of personal data is a naturally occurring phenomenon - the essence of the principle of data accuracy, however, is that the controller strives, within the framework of all reasonable measures, to correct or rectify erroneous data. According to the accepted position in the doctrine[22]:

> *The principle of data accuracy, however, should not be interpreted as an obligation imposed on the controller to systematically search for inaccurate data. In practice, such an approach would be extremely difficult to implement, not only because of the volume of data processed, but also because of the problems of verifying its correctness. Therefore, the commented provision of the regulation imposes an obligation on the controller to update the data "when necessary." This means that the administrator should respond to signals of irregularities* (...)

Data accuracy should be ensured, in particular, if the data subject requests rectification of their personal data in accordance with Article 16 GDPR. This provision is a practical form of implementing the principle of data accuracy.

In the case of OpenAI and the processing of data to train models, this principle is **completely ignored** in practice. This is evidenced by OpenAI's response to Mr. Łukasz Olejnik's request, according to which OpenAI was unable to correct the processed data. OpenAI's systemic inability to correct data is assumed by OpenAI as part of ChatGPT's operating model.

In accordance with the Privacy Policy (machine translation):

> *Note on accuracy: Services such as ChatGPT generate responses by reading the 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 accurate. For this reason, you should not rely on the actual accuracy of the output from our models. If you notice that ChatGPT's output contains inaccurate information about you and want us to correct the inaccuracy, you can send a correction request to dsar@openai.com. Given the technical complexity of how our models work,* ***we may not be able to correct inaccuracies in every case****. If this is the case, you may request that we remove your Personal Information from ChatGPT's output by filling out this form.*

Given the general and vague description of ChatGPT's data validity mechanisms, it is highly likely that the inability to correct data is **a systemic phenomenon** in OpenAI's data processing, and not just in limited cases.

The above circumstances may raise reasonable doubts about the overall compliance with data protection regulations of a tool, an essential element of which is the systemic inaccuracy of the processed data. These doubts are reinforced by the scale of ChatGPT's processed data and the scale of potential recipients of personal data, which affect the risks to rights and freedoms associated with personal data inaccuracy.

Also relevant to the evaluation of the issues described is the fact that the processing of personal data within ChatGPT is not just about presenting the source data with which the models are fed (as, for example, is the case with classic search engines). The essence of the inaccuracy of personal data processed by OpenAI is not only the inaccuracy of the source data. In the case of ChatGPT-type tools, data inaccuracy also results from the processing of data (as part of machine learning) to generate content, and refers in this case to **"output" data.** The risk of this inaccuracy

---

[21] P. Fajgielski [in:] Commentary to Regulation 2016/679 on the protection of individuals with regard to the processing of personal data and on the free movement of such data and repealing Directive 95/46/EC (General Data Protection Regulation) [in:] General Data Protection Regulation. Personal Data Protection Law. Commentary, 2nd edition, Warsaw 2022, Article 5.

[22] Ibid.

is also compounded by the way ChatGPT provides the output data, which is often presented as fact, regardless of the accuracy of the data.

Therefore, in order to comply with its obligations under Article 5(1)(d) GDPR and Article 16 GDPR, OpenAI should strive to correct errors created by the generation of content in response to users' questions. OpenAI should develop and implement a data rectification mechanism based on an appropriate filter/module that would verify and correct content generated by ChatGPT (e.g., based on a database of corrected results). It is reasonable in the context of the scope of the obligation to ensure data accuracy to expect OpenAI to correct at least data reported or flagged by users as incorrect. Such situation occurred in the case of Mr. Łukasz Olejnik's request.

We believe that it is possible for OpenAI to develop adequate and GDPR-compliant mechanisms for correcting inaccurate data (it is already possible to block the generation of certain content as a result of a blockade imposed by OpenAI). However, if, in OpenAI's opinion, it is not possible to develop such mechanisms - it would be necessary to consult the issue with the relevant supervisory authorities, including, for example, through the prior consultation procedure described in Article 36 of GDPR.

Conclusions. In the opinion of Mr. Łukasz Olejnik, there was a violation of his right to rectification of data - this right was not executed, and instead, contrary to the request, there was a restriction of data processing.

Moreover, from the information provided in the email correspondence, as well as from the information available online, it appears that OpenAI's inability to correct data (exercise the rights of individuals) is systemic, causing OpenAI's data processing to violate the principle of data accuracy contained in Article 5(1)(d) GDPR.

2.5.C. Other violations related to improper execution of Mr. Łukasz Olejnik's rights

Regardless of the violations in the execution of Mr. Łukasz Olejnik's specific rights described in sections 2.3.1 and 2.3.2 - also the general manner in which these rights are exercised indicates processing that does not comply with Article 12 of GDPR.

OpenAI's model of responding to data subject requests demonstrates **a facade exercise of rights**, contrary to OpenAI's obligations as a data controller. This means that OpenAI acts in a way intended to give the impression of acting in accordance with the GDPR (in terms of exercising the rights of the individual), but in reality it exercises these rights in violation of GDPR.

In accordance with Article 12 (1) and (2) of the GDPR:

> *(1) The controller shall take appropriate measures to provide any information referred to in Articles 13 and 14 and any communication under Articles 15 to 22 and 34 relating to processing to the data subject in a concise, transparent, intelligible and easily accessible form, using clear and plain language, in particular for any information addressed specifically to a child.* (...)
>
> *(2) The controller shall facilitate the exercise of the data subject's rights under Articles 15 to 22.*
>
> (...)

In the correspondence exchanged by Mr. Łukasz Olejnik with OpenAI, the controller provided a variety of information, some of it unrelated to Mr. Łukasz Olejnik's request. Moreover, the set of information provided (despite its comprehensiveness) did not include specific information in accordance with the scope of Article 15(1) GDPR. The exercise of the right also did not take place despite the fact that Mr. Łukasz Olejnik sent more emails, in part to clarify the request, inquiring about specific issues, indicating that OpenAI's responses did not address the issues covered by the request.

In addition, the content of OpenAI's response and the manner in which certain information was provided is questionable. OpenAI's clarification contained numerous links, references to other pages in OpenAI's web domain. The reference included the issue of OpenAI's processing of data to train models - it can be argued that OpenAI is obstructing access to information in this regard. Moreover, actually bringing about the exercise of rights (obtaining specific information) would often require the user to direct additional questions, depending on the issue, through different channels

and to different addressees. Correspondence with OpenAI may also identify misrepresented information (e.g., an initial statement that automated decision-making is taking place, or indicating a list of sub-processors in response to a question about recipients of personal data).

This way of exercising OpenAI's right of access does not comply with the obligation to exercise the data subject's rights **in a concise, let alone clear and understandable form**. OpenAI's action also violates, in Mr. Łukasz Olejnik's opinion, Article 12(2) GDPR, which obligates the controller to facilitate the data subject's exercise of his or her rights.

The literature[23] indicates that:

> *Among the guiding principles for data processing introduced by the provisions of the General Regulation is the principle of transparency, which is enshrined in Article 5(1)(a) of the act. The commented article* [Article 12 of the GDPR] *refers it to the mode of exercise of the rights of data subjects set forth in Chapter III, setting a consistent framework of conduct for all processors covered by the provisions of the General Regulation.*
>
> (...)
>
> *The legitimacy of adopting the principle of transparency seems unquestionable. In fact, it is an expression of the concern that the European legislator attaches to empowering the **data subject** by ensuring the due exercise of his or her rights, including those of an informative nature.* (...) *Communications from the controller are therefore intended not only to be clear to the data subject, but also to serve as a kind of **guide**.*

Conclusion. Therefore, the described modus operandi of OpenAI in exercising the rights of Mr. Łukasz Olejnik violated Article 12(1) and (2) GDPR as well as the transparency principle in Article 5(1)(a) of the GDPR.

2.6. Violation of data protection by design (privacy by design) principles

Violation of the principle of data protection by design. In the opinion of Mr. Łukasz Olejnik, OpenAI completely **ignored the principle of data protection by design when** creating and making Chat-GPT available.

The principle of data protection by design is described in Article 25(1) of the GDPR, according to which:

> *Taking into account the state of the art, the cost of implementation and the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for rights and freedoms of natural persons posed by the processing, the controller shall, both at the time of the determination of the means for processing and at the time of the processing itself, implement appropriate technical and organisational measures, such as pseudonymisation, which are **designed to implement data-protection principles, such as data minimisation**, in an effective manner and to integrate the necessary safeguards into the processing in order to meet the requirements of this Regulation and protect the rights of data subjects.*

Data protection by design (or privacy by design) is a concept at the heart of which is the consideration of privacy in the design phase. The basic tenets of this principle are in particular[24]:

1) ***A proactive, not reactive, approach**; preventive, not corrective. The PbD approach anticipates privacy-invasive events before they happen and prevents them.* (...)
2) ***Privacy as a default value**. The PbD concept is based on the idea that we can all be sure of one thing - the default rules. Privacy by Design seeks to provide the maximum degree of*

---

[23] J. Luczak [in:] RODO. General Data Protection Regulation. Commentary, ed. by E. Bielak-Jomaa, D. Lubasz, Warsaw 2018, Article 12.

[24] Based on: P. Fajgielski [in:] Commentary to Regulation No. 2016/679 on the protection of individuals with regard to the processing of personal data and on the free movement of such data and repealing Directive 95/46/EC (General Data Protection Regulation) [in:] General Data Protection Regulation. Personal Data Protection Law. Commentary, 2nd edition, Warsaw 2022, Article 25 and the source material cited therein: Privacy by Design, The 7 Foundational Principles, Ann Cavoukian, Ph.D., https://www.ipc.on.ca/wp-content/uploads/Resources/7foundationalprinciples.pdf (accessed 04.08.2023).

privacy protection by ensuring that personal data is automatically protected in any information system or business practice. If a person does nothing, their privacy will still be intact. (...)

3) ***Privacy built into the design****. The concept of PbD is built into the design and architecture of information systems and business practices. It is not included as an add-on, after the fact.* (...)

4) ***Security from start to finish*** *- protection throughout the information lifecycle. The PbD concept built into the system before the first information is collected breaks down securely over the entire life cycle of the data in question - robust security measures are essential to protect privacy, from start to finish.* (...)

5) ***Visibility and transparency****. The PbD concept seeks to reassure all stakeholders that, regardless of the business practice or technology used, it is in fact operating according to its stated promises and goals, subject to independent verification. Its components and operations remain visible and transparent, equally to users and suppliers.* (...)

6) ***Respecting user privacy****. Above all, the PbD concept requires that architects and operators prioritize a person's interests, offering measures such as robust privacy defaults and adequate notice, and providing user-friendly options. The user is to be the focus of*[25] *.*

The way the ChatGPT tool was designed, taking into account also the violations described in paragraph 3.3 of the complaint (in particular, the inability to exercise the right to rectify data, the omission of data processing operations for training GPT models) - **contradicts all the indicated assumptions of the principle of data protection by design**. In practice, in the case of data processing by OpenAI, there is testing of the ChatGPT tool using personal data, not in the design phase, but in the production environment (i.e., after the tool is made available to users).

OpenAI seems to accept that the ChatGPT tool model that has been developed is simply incompatible with the provisions of GDPR, and it agrees to this state of affairs. This shows a complete disregard for the goals behind the principle of data protection by design.

3. Concluding remarks

The factual and legal circumstances described in this complaint confirm that there was unlawful processing of Mr. Łukasz Olejnik's personal data by OpenAI. OpenAI's action violated:

- **Article 5(1)(a) GDPR** - OpenAI processed Mr. Łukasz Olejnik's data unlawfully, unfairly, and in a non-transparent manner,
- **Articles 12 and 15 GDPR** - OpenAI improperly exercised Mr. Łukasz Olejnik's right to access personal data and to information about the processing of personal data,
- **Articles 12 and 16 GDPR -** OpenAI has not exercised Mr. Łukasz Olejnik's right to rectify his inaccurate data,
- **Article 25(1) GDPR** - OpenAI, in designing and implementing the ChatGPT tool, violated the principle of data protection by design.

[25] This also means that one should design taking into account the possibility of exercising the rights of individuals in connection with the processing of personal data, M. Gawroński (ed.) Personal Data Protection. Guide to the Act and RODO with models, Warsaw 2018, pp. 341-342.

In view of these violations, it is reasonable to oblige OpenAI to exercise, in accordance with Articles 12, 15 and 16 GDPR, the rights of Mr. Łukasz Olejnik, i.e. the right of access to personal data and information about the processing of personal data, and the right to rectify personal data.

It would also be expedient for the authority, given the above violations, most likely systemic in nature, to require OpenAI to submit a Data Protection Impact Assessment (DPIA) document. This document could be an important element in assessing whether OpenAI's processing of data within the Chat-GPT tools is in compliance with the GDPR.

New technological solutions, especially tools as innovative and technologically complex as generative artificial intelligence tools with universal applicability and widespread access, and ChatGPT is such a tool - should be developed taking into account the models of *right-based approach* (an approach based on the individual rights of data subjects) and *risk-based approach* (an approach based on the risks that data processing entails, in particular to individual rights and freedoms)[26] . As the UK DPA rightly pointed out:

> *Organizations developing or using generative artificial intelligence should consider their data protection obligations from the outset, adopting a privacy by design and privacy by default approach. This isn't optional - if you process personal data, it's the law[27].*

For the above reasons, I conclude as in the introduction.

_____

Maciej Gawronski
Attorney-at-law

Attachments:

(1) A copy of the power of attorney for attorney-at-law Maciej Gawronski,

(2) Proof of payment on the power of attorney,

(3) Screenshot #1 containing an excerpt from a conversation between Mr. Łukasz Olejnik and ChatGPT,

(4) Screenshot #2 containing an excerpt from a conversation between Mr. Łukasz Olejnik and ChatGPT,

---

[26] The impact of the General Data Protection Regulation (GDPR) on artificial intelligence, Study of the Panel for the Future of Science and Technology, European Parliamentary Research Service, June 2020, https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU(2020)641530_EN.pdf (accessed 04.08.2023).

[27] Generative AI: eight questions that developers and users need to ask, own translation, https://ico.org.uk/about-the-ico/media-centre/news-and-blogs/2023/04/generative-ai-eight-questions-that-developers-and-users-need-to-ask/, (accessed August 01, 2023).

(5) Email correspondence between Mr. Łukasz Olejnik and OpenAI regarding the processing of personal data within ChatGPT.

(6) Machine translation of Mr. Łukasz Olejnik's e-mail correspondence with OpenAI regarding the processing of personal data within ChatGPT.