Complaint filed in accordance with Article 58 of the Act of 3 December 2017

(Informal machine translation)

Filed by:

Pierre Dewitte, residing at ████████████████████████████
You can reach us by e-mail at pierre.dewitte@mailfence.com

And

Geertrui Mieke De Ketelaere, residing at ███████████████████████
You can reach us by e-mail at miekedk@hotmail.com

Against:

Chai Research Corp. with registered office at 39899 Balentine Drive, Suite 200, Newark, California, 94560, United States

---

1. Introductory remarks

1.1. As to the use of an hoc hoc form rather than the one provided by the DPA

This complaint is submitted by means of an *ad hoc* document, rather than through the form available on the Data Protection Authority (hereinafter "DPA")'s website,[1] insofar as the latter imposes space and structural restrictions that do not allow a clear and complete description of the relevant facts.

1.2. As to the territorial applicability of the GDPR and the Law of 30 July 2018

In accordance with Article 3(2), a and b of the General Data Protection Regulation (hereinafter "GDPR"), transposed respectively into Belgian law by Article 4, § 2, 1° and 2° of the Law of 30 July 2018 on the protection of individuals with regard to the processing of personal data (hereinafter "Law of 30 July 2018"), these two texts apply to Chai Research Corp. insofar as, although the company does not have a relevant establishment on Belgian territory, the processing operations referred to in point 3 are nevertheless linked to the provision of services to data subjects on Belgian territory, and involve the monitoring of their behaviour taking place on Belgian territory.

1.3. As to the competence of the Data Protection Authority

Since, as pointed out in point 1.2, Chai Research Corp. does not have a relevant establishment on the Belgian territory or in another EU country, the processing operations referred to in point 3 cannot be qualified as "*cross-border processing*" within the meaning of Article 4(23) of the GDPR. Consequently, the cooperation mechanism established by Articles 56 and 60 of the GDPR does not apply. There is therefore no need to identify the lead supervisory authority or the supervisory authorities concerned.

---

[1] The dedicated form made available by the APD can be accessed at
https://www.autoriteprotectiondonnees.be/publications/formulaire-pour-introduire-une-plainte.pdf

The DPA is therefore competent to decide on the processing operations referred to in point 3 on Belgian territory by virtue of Article 4, §1er , paragraphs 1 and 2 of the Law of 3 December 2017 establishing the Data Protection Authority (hereinafter "Law of 3 December 2017").

1.4. As to the subsequent correspondence between the APD and the complainants

We kindly request the DPA to send all further communications related to the handling of this complaint to the two complainants via the contact details mentioned in the header.

2. Overview of how the "Chai" application works

The "Chai" application available on Android via the Google Play Store[2] and on iOS via the Apple App Store,[3] is presented as a platform for easily developing and deploying conversational robots (hereafter "chatbots").[4] In concrete terms, its operation is based on two elements: the creation of personalised chatbots (point 2.1) and the provision of these chatbots to other users of the platform (point 2.2). This is made possible thanks to a complex architecture involving several actors (point 2.3).

2.1. The creation of custom chatbots

Concerning the first aspect, users have the possibility, after creating an account and identifying themselves on the platform, to customise a "blank" chatbot on the basis of defined criteria in order to influence its behaviour and responses. They have two options to do this. Either manually enter this information in the parameters of the bot in question. Or use "ChaiPy", the Python interface available to developers.[5] The customisation criteria include, among others, the following:

- The name of the chatbot.
- A general description to identify the chatbot.
- A 'welcome message' which serves as a basis for starting all conversations with the user. This message can also suggest a general atmosphere and role-playing elements if the user wishes.
- A "memory" that serves as a starting point for the personalisation of the chatbot and contains information such as its communication style, its relationship with and attitude towards the user, as well as all the details that will persist throughout the various conversations.
- A "prompt" that establishes the general context according to which the chatbot will orient its answers. This prompt can take the form of an example of the type of conversation the user would like to have with the chatbot, or a description of the scene that will serve as a backdrop to the conversation between the chatbot and the user.
- Other optional parameters such as the "temperature" (*i.e.* the randomness of the answers provided by the chatbot), the "repetition penalty" (*i.e.* the degree of repetition allowed) or the "chatbot label" (*i.e.* the way the chatbot refers to itself).

---

[2] The Android application can be found at https://play.google.com/store/search?q=chai&c=apps.

[3] The iOS application can be found at https://apps.apple.com/be/app/chai/id1544750895.

[4] Information about the "Chai" application is scattered across several sources. These include the Chai Research Corp. GitHub repository at https://github.com/chai-research, the HuggingFace page for the "Chai" application at https://huggingface.co/ChaiML, and the information in the documentation on the official "Chai" website at https://chai.ml/docs. The Chai sub-forum on Reddit (https://www.reddit.com/r/ChaiApp/) also contains a wealth of user-generated information. For example, see the "How To: Chai 101" guide at https://www.reddit.com/r/ChaiApp/comments/11rp9gp/how_to_chai_101/ and the "A Quick Guide to Chai AI: Building a Bot!" guide on the "Replika" SubReddit at https://www.reddit.com/r/replika/comments/111cdc5/a_quick_guide_to_chai_ai_building_a_bot/.

[5] The Python interface is available at https://github.com/chai-research/chai_py.

A YouTube video illustrates this customisation process,[6] as well as Figures 1 to 3.

| Figure 1 - Bot 1 customization window | Figure 2 - Bot 2 customization window | Figure 3 - Bot customization window 3 |

These customisation elements can be succinct, or contain long descriptions. The more precise they are, the more specific the chatbot's behaviour will be. In any case, the parameters defined by the user are only starting points in the chatbot's customisation. The chatbot will evolve as it talks to the user, and will adjust its responses and behaviour according to the user's reactions. The chatbot will also remember information shared by the user in previous conversations.

2.2. The provision of chatbots to other users of the platform

Once personalised, users are free to make their chatbots available to other users on the platform. Chai Research Corp. encourages this sharing by maintaining a ranking of the most popular chatbots based on user engagement and average conversation length.[7] Exhibit 1 provides an overview of the ranking as of April 13, 2023.

---

[6] The video is available at https://www.youtube.com/watch?v=BWr_CJQc6pU.
[7] The ranking can be found at https://chai.ml/dev.

2.3. How the "Chai" application works

The "Chai" application is based on the use of the "GPT-J 6B" language model developed by the company EleutherAI.[8] This model was itself trained on the basis of "The Pile", an open source dataset of more than 825 gigabytes comprising 22 smaller datasets for training language models.[9] These 22 datasets are themselves derived from publicly available sources such as PubMed Central, ArXiv, GitHub, the FreeLaw Project, Stack Exchange, the US Patent and Trademark Office, PubMed, Ubuntu IRC, HackerNews, YouTube, PhilPapers, and NIH ExPorter, combined with existing datasets such as Books3, Project Gutenberg, Open-Subtitles, English Wikipedia, DM Mathematics, EuroParl, and the Enron Emails corpus.[10] We also draw the attention of the ODA to the fact that "GPT-J 6B" has a high degree of bias and toxicity, as demonstrated in a study of April 2023 (figures 4 and 5).[11]

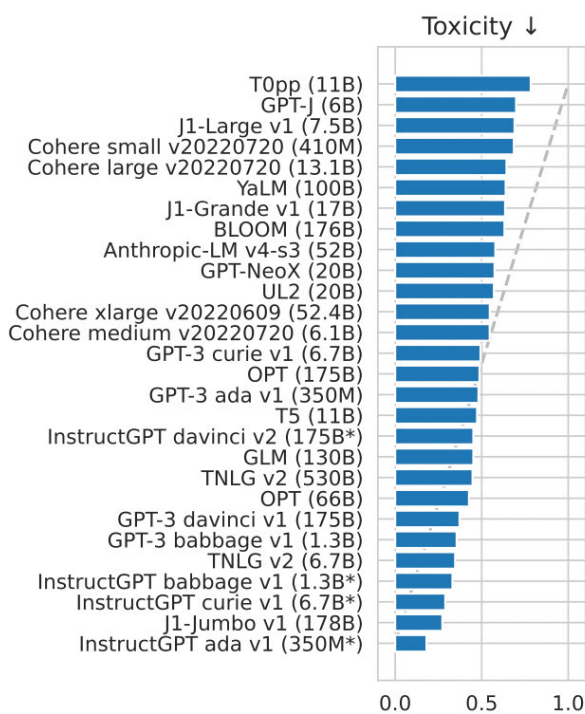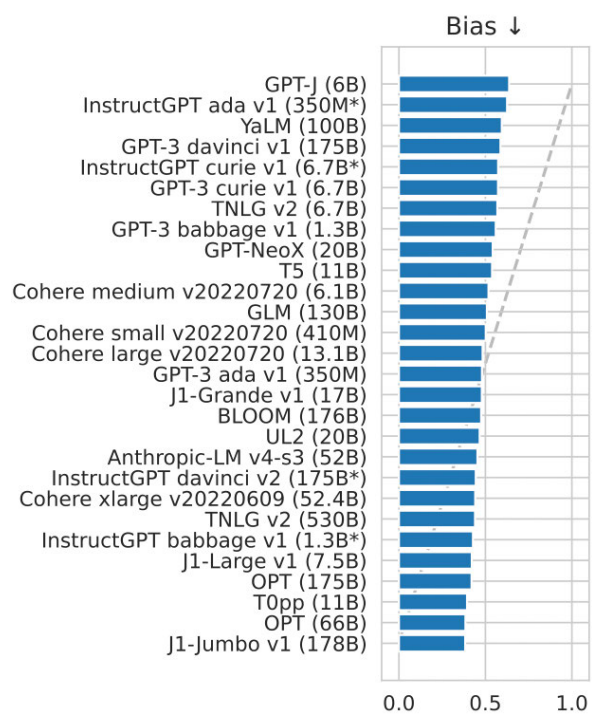Figure 4 - Toxicity level of "GPT-J 6B

Figure 5 - Toxicity level of "GPT-J 6B

The "Chai" application sits between "GPT-J 6B" and users, allowing them to personalise their own chatbots through the elements described in section 2.1. To do this, it would appear that each message sent by the user to a personalised chatbot is accompanied by the personalisation parameters they have set.[12] By combining this information with previous conversations and information shared by the users themselves, the chatbots offered on the platform are likely to evolve freely. However, the exact functioning of the application is difficult to observe, as it is not open source, and the "desktop" version that allowed interaction with the chatbots from a web browser was recently suspended.

---

[8] The HuggingFace page of "GPT-J 6B" can be found at https://huggingface.co/EleutherAI/gpt-j-6b.

[9] The Pile dataset can be accessed at https://pile.eleuther.ai/.

[10] Leo Gao and others, 'The Pile: An 800GB Dataset of Diverse Text for Language Modeling' (arXiv, 31 December 2020) http://arxiv.org/abs/2101.00027 accessed 12 April 2023.

[11] Cem Dilmegani, 'The Future of Large Language Models' (*AI Multiple*, 10 April 2023) https://research.aimultiple.com/future-of-large-language-models/ accessed 15 April 2023.

[12] See in particular the blog post available at https://chai.ml/blog/gpt-j/.

It appears from the previous paragraphs that the functioning of the "Chai" application involves several actors. As detailed in point 3, it is also possible to distinguish different groups of processing activities involving personal data. The following sections detail, for each group of processing activities with a similar purpose, the controller(s) involved and the infringements of the GDPR attributable to them.

3. The processing involved and the infringements identified

3.1. Training of the GPT-J language model by EleutherAI

3.1.1. The processing of personal data at issue

As mentioned in section 2.3, the "Chai" application is largely based on the "GPT-J 6B" language model developed by the company EleutherAI. In view of the diversity of the dataset "The Pile" used for training this model (see figure 6), it inevitably contains personal data within the meaning of Article 4(1) of the GDPR. Without being exhaustive, this is particularly the case of information shared by users on platforms such as GitHub, Wikipedia or StackExchange, as well as contact and affiliation information referenced in scientific articles published on PubMed Central or ArXiv. These data are collected "en masse" for re-use in the context of the "GPT-J 6B" training, without any distinction or sorting between personal data and data that do not relate to identified or identifiable individuals.
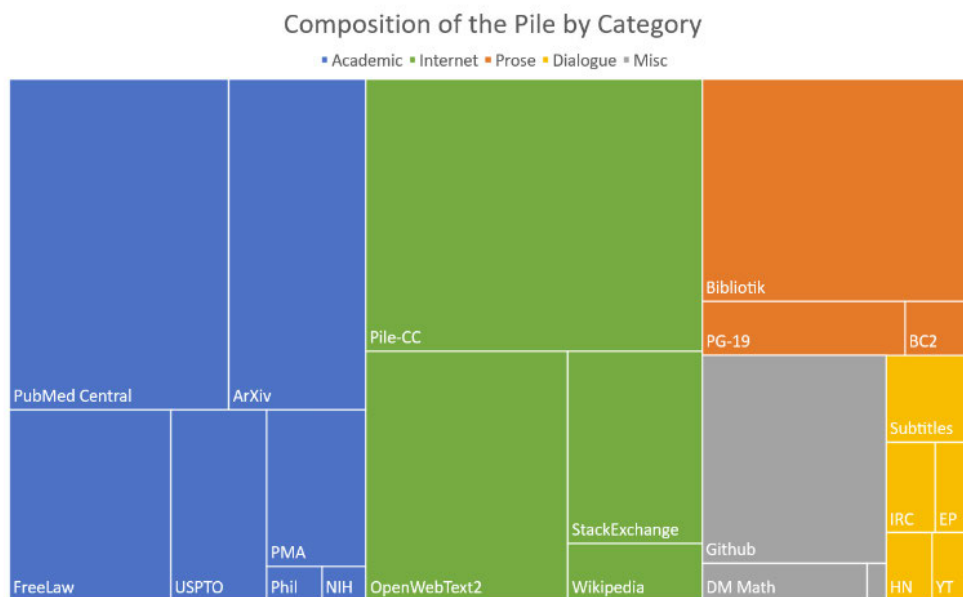

Figure 6 - Contents of 'The Pile' dataset, from Gao L and others, 'The Pile: An 800GB Dataset of Diverse Text for Language Modeling' (arXiv, 31 December 2020) <http://arxiv.org/abs/2101.00027> accessed 12 April 2023 (p. 2).

3.1.2. The controller(s)

EleutherAI acts as a controller within the meaning of Article 4(7) of the GDPR with regard to the processing of personal data contained in the dataset "The Pile" insofar as it determines the purpose (*i.e.* the training of its language model "GPT-J 6B") and the means (*i.e.* in particular the training procedure).[13]

---

[13] According to EleutherAI's HuggingFace page, "GPT-J 6B" has been trained for 402 billion tokens on 383,500 steps on TPU v3-256 pod. It was trained as an autoregressive language model, using cross-entropy loss to maximise the probability of correctly predicting the next word.

3.1.3. Infringements identified

Although this complaint is not directed against EleutherAI, we would like to draw the attention of the DPA to the fact that the company has no legal basis to justify the massive downloading of data publicly available on the internet and their reuse for the purpose of training artificial intelligence systems of the "machine learning" type. The processing operations referred to in point 3.1.1. are therefore in breach of the lawfulness principle laid down in Articles 5(1)a and 6(1) of the GDPR, insofar as EleutherAI has not obtained the consent of the data subjects, nor does it invoke how its legitimate interests could outweigh those of the data subjects. There is also no contract to justify the processing in question.

This was confirmed by the Commission Nationale de l'Informatique et des Libertés (hereinafter: "CNIL") in its decision MED-2021-134 against the company Clearview AI for having developed facial recognition software based on "*the systematic and widespread collection, from millions of websites worldwide, of images containing faces, using a proprietary technology to index freely accessible web pages*".[14] This decision was fully followed by its Italian and Greek counterparts, both of which also fined Clearview AI for failing to comply with the lawfulness requirement in relation to the mass and widespread collection of photographs for training purposes.[15]

More recently, and in a context similar to the one that serves as a backdrop to the present complaint, the Garante per la protezione dei dati personali (hereinafter: "Garante") prohibited the use of the "ChatGPT" service developed by the company OpenAI on Italian territory, citing, among other failings, the lack of "*an appropriate legal basis for the collection of personal data and their processing for the purpose of training the algorithms underlying the operation of ChatGPT*".[16] On 12 April 2023, the Garante also required the company to "*remove all references to contractual performance and to rely – in accordance with the principle of accountability – on consent or legitimate interest as the legal basis*".[17] A similar complaint was recently submitted to the CNIL.[18]

While the breach of the principle of lawfulness is the most fundamental violation, the practices referred to in point 3.1.1 are also accompanied by other breaches such as a lack of transparency and breaches of obligations arising from the exercise of the data subject's rights. These do not constitute an exhaustive list, and we do not exclude that the DPA adds others to the list.

In view of the above, we ask the DPA, through its Inspection Service, to investigate the practice of massive scraping of publicly available data from the internet for subsequent re-use in training machine learning tools, and to assess their compliance with the requirements set out in the GDPR. These models, once trained, are effectively used as a basis for the development and deployment of many services – such as the "Chai" application, which is de *facto* based on previous illegal data collection. It is therefore crucial to address the root of the problem.

---

[14] Commission Nationale de l'Informatique et des Libertés, decision MED-2021-134 of 26 November 2021, available at https://www.legifrance.gouv.fr/cnil/id/CNILTEXT000044499030.

[15] Garante per la protezione dei dati personali, Ordinanza ingiunzione nei confronti di Clearview AI - 10 febbraio 2022 [9751362], available at https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9751362; Αρχή προστασίας δεδομένων προσωπικού χαρακτήρα, Επιβολή προστίμου στην εταιρεία Clearview AI, Inc. available at https://www.dpa.gr/el/enimerwtiko/prakseisArxis/epiboli-prostimoy-stin-etaireia-clearview-ai-inc.

[16] Garante per la protezione dei dati personali, Provvedimento del 30 marzo 2023 [9870832], available at https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9870832.

[17] Garante per la protezione dei dati personali, Provvedimento dell'11 aprile 2023 [9874702], available at https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9874702.

[18] David Libeau, 'I Filed A Complaint To The CNIL Against ChatGPT' (*blog.DavidLibeau.fr*, 2 April 2023) https://blog.davidlibeau.fr/jai-depose-une-plainte-a-la-cnil-contre-chatgpt/ accessed 13 April 2023.

3.2. Processing of user data by Chai Research Corp. for service purposes

3.2.1. *The personal data processing at issue*

The use of the "Chai" application requires the collection and processing of several types of personal data. Firstly, account-related data, the creation of which is necessary in order to be able to customise one's own chatbot or use those made available by users. This data includes information related to the third party account used by the user to open his "Chai" account, such as Twitter, Facebook, Google or Apple.[19] This includes, but is not limited to, the name as entered in the account used, profile picture, email address, and other information depending on the third party social network used to create the "Chai" account, as described in the "How do we handle your social logins?" section of the application's privacy policy.[20] Exhibit 2 reproduces the Chai privacy policy and Exhibit 3 provides a French machine translation. Figures 7 and 8 illustrate some of the personal data collected during account creation. The application also creates a "User ID" automatically.
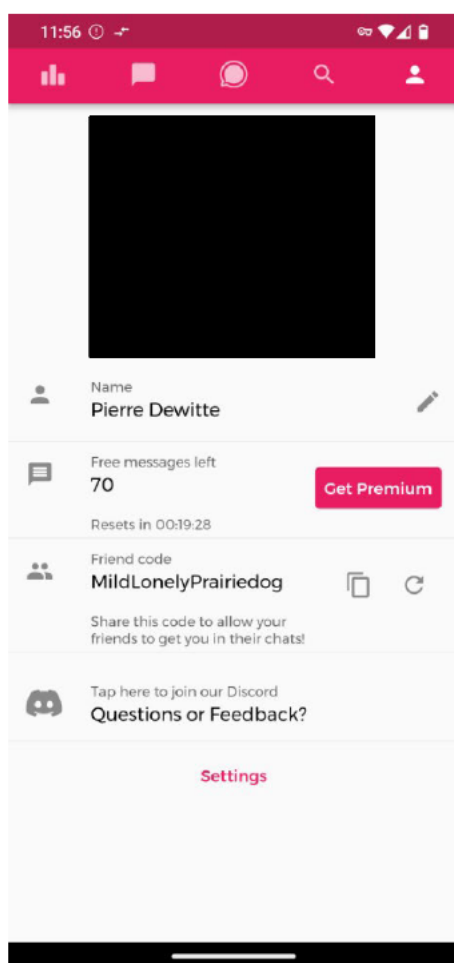

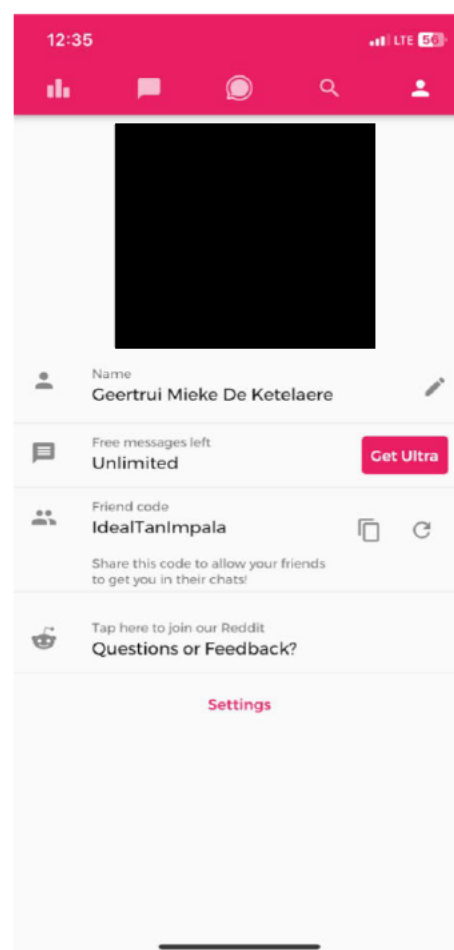Figure 7 - Data associated with Pierre Dewitte's account

Figure 8 - Data associated with the Geertrui Mieke De Ketelaere account

---

[19] For an overview of the account creation process, see the YouTube video available at https://www.youtube.com/watch?v=w4XdPYOhwsI.

[20] It can be accessed at https://chai.ml/privacy/.

Secondly, the application collects a range of usage personal data. This includes, as stated in the "Information automatically collected" and "Information collected through our App" sections of the Chai Privacy Policy, "*IP address, browser and device characteristics, operating system, language preferences, a referring URL, device name, country, location, information about how and when you use our services, and other technical information*. With respect to the mobile application in particular, this data includes "*device information (such as the ID, model and manufacturer of your mobile device), operating system, system version and configuration information, device and application identification numbers, browser type and version, hardware model, Internet service provider and/or mobile operator, and Internet Protocol (IP) address (or proxy server)*". The Privacy Policy also states that "*if you use our application, [Chai may] also collect information about the telephone network associated with your mobile device, the operating system or platform of your mobile device, the type of mobile device you are using, the unique identifier of your mobile device, and information about the features of our application that you have accessed*. According to the privacy policy, this data is "*primarily necessary to maintain the security and operation of our services, as well as for internal analysis and reporting purposes*" and "*to maintain the security and operation of our application, for troubleshooting and for internal *analysis and reporting purposes*. It is important to note at the outset, and contrary to what Chai seems to suggest in its privacy policy when the company states that "*this information does not reveal your specific identity (such as your name or contact details)*", that this does constitute personal data within the meaning of Article 4(1) of the GDPR.

Finally, the application also processes the personal data needed to personalise the chatbots. This data includes in particular the "welcome message", the "memory" and the "prompt" as described in point 2.1. if the user decides to personalise a "blank" chatbot. This data also includes, and in all cases, the content of the conversations held between the user and the chatbots in order to refine their functioning. It is important to note that, although this information is to be considered as personal data since it is associated with the other information described above (notably account information and automatically collected identifiers), its content may also reveal details that allow the user to be identified. This is the case, for example, when the user voluntarily shares information about himself in the context of his conversations with chatbots, or in one of the elements of personalisation mentioned in point 2.1. Figures 9 and 10 illustrate these assumptions. Chai's privacy policy makes no mention of such processing nor of such data.

As explained in section 2.3, the exact functioning of the "Chai" application is difficult to observe. Nevertheless, the associated privacy policy does not mention all the processing operations involved in using the service. We therefore ask the DPA, through its Inspection Service, to examine all processing activities and personal data collected and processed in the operation of "Chai". A clear overview of (i) the processing operations in question, (ii) the associated data and (iii) their respective purposes is indeed a prerequisite for examining the compliance of the "Chai" application with the GDPR. This information should also be clearly reflected in the register kept in accordance with Article 30(1) of the GDPR, a copy of which we request the DPA to obtain as provided for in Article 30(4) of the GDPR. The breaches detailed in section 3.2.3 are therefore based on our own observations, in addition to the processing mentioned in the "Chai" privacy policy.
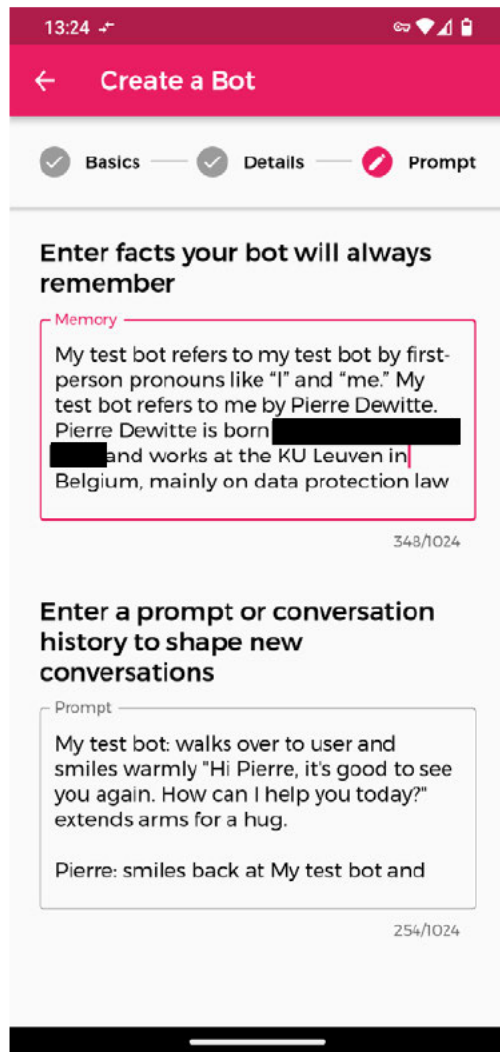
Figure 9 - Personal data contained in the "memory" and "prompt"

Figure 10 - Personal data exchanged in the conversation

3.2.2. The controller(s)

There are several possible readings of the allocation of responsibilities in the case at stake. According to the first, Chai Research Corp. acts, alone, as a controller within the meaning of Article 4(7) of the GDPR, insofar as it determines both the purposes and the means of the processing activities detailed in point 3.2.1 (*i.e.* the processing operations whose purpose is to offer the "Chai" service to its users). Chai Research Corp. also seems to implicitly acknowledge this qualification when it tries to comply, through its privacy policy, with an obligation of transparency that is only incumbent on data controllers (as detailed in point 3.2.3, sub-point b).

It is indeed realistic to argue that Chai Research Corp. determines the purposes for which the data listed in point 3.2.1 are collected and processed (*i.e.* in order to provide a platform for customising, deploying and using chatbots). It is indeed Chai Research Corp. that decides to develop and make available an application performing the functions described in point 2. It is also this company that determines both the "essential means" (*i.e.* It is also this company that determines both the "essential" (i.e. "*the type of personal data that are processed, the duration of the processing, the categories of recipients as well as the categories of data subjects*") and "non-essential" (*i.e.* "*the more practical aspects of the implementation, such as the choice of a particular type of hardware or software or the*

detailed security measures") means through which these activities are carried out[21] As highlighted in section 2.3, it is indeed Chai Research Corp that has implemented the interface and functionalities that allows users to customise and use their chatbots.

This first reading also implies, in the specific context of the processing operations referred to in point 3.2.1, that EleutherAI can be considered as a subcontractor of Chai Research Corp. since the company limits itself to processing the data of users of the "Chai" application on behalf of Chai Research Corp. While it is indeed the "GPT-J 6B" language model developed by EleutherAI that is used to generate the chatbot interactions accessible on the "Chai" application, it is Chai Research Corp. that provides the technical solutions that ensure the degree of personalisation expected by users. However, it is the provision of the personalised chatbots service, and not a neutral language model, that is the purpose of the processing activities described in point 3.2.1. In that sense, EleutherAI only plays an instrumental role in the provision of this service. This reading of the notion of "controller" and "processor" finds a semblance of echo in the guidance proposed by the Information Commissioner's Office, which illustrates the aspects of responsibility in the field of artificial intelligence by means of the example shown in figure 11.[22] It is possible to reason by analogy and draw similar conclusions with regard to the relationship between Chai Research Corp. and EleutherAI.

💡 **Example**

An organisation provides live AI prediction and classification services to clients. It develops its own AI models, and allows clients to send queries via an API ('what objects are in this image?) to get responses (a classification of objects in the image).

First, the prediction service provider decides how to create and train the model that powers its services, and processes data for these purposes. It is likely to be a controller for this element of the processing.

Second, the provider processes data to make predictions and classifications about particular examples for each client. The client is more likely to be the controller for this element of the processing, and the provider is likely to be a processor.

Figure 11 - Example from the ICO guidance on AI and the GDPR

A second reading, in which Chai Research Corp. and EleutherAI act as joint controllers for the operations referred to in point 3.2.1, is also possible. This view follows, inter alia, from the criteria put forward by the Court of Justice of the European Union (hereinafter "CJEU") in cases C-210/16,[23] C-25/17[24] and C-40/17,[25] as well as by the European Data Protection Board (hereinafter "EDPB") in its Guidelines 07/2020.[26] In the above-mentioned cases, the CJEU has consistently stressed the need to "*ensure, through a broad definition of the concept of 'controller', effective and comprehensive*

---

[21] European Data Protection Board, "Guidelines 07/2020 on the concepts of controller and processor in the GDPR", especially paragraphs 30-39 regarding purposes and means, available at https://edpb.europa.eu/sites/default/files/consultation/edpb_guidelines_202007_controllerprocessor_en.pdf.
[22] The guidance is available at https://ico.org.uk/for-organisations/guide-to-data-protection/key-dp-themes/guidance-on-ai-and-data-protection/what-are-the-accountability-and-governance-implications-of-ai/#howshouldweunderstand.
[23] Case C-210/16, *Wirtschaftsakademie Schleswig-Holstein GmbH*, 5 June 2018, ECLI:EU:C:2018:388.
[24] Case C-25/17, *Tietosuojavaltuutettu*, 10 July 2017, ECLI:EU:C:2018:551.
[25] Case C-40/17, *Fashion ID GmbH & Co* KG, 29 July 2019, ECLI:EU:C:2019:629.
[26] See note 21.

protection of data subjects" (see C-210/16, paragraph 28; C-25/17, paragraph 66 and C-40/17, paragraph 66). More specifically, this reasoning can be justified as follows.

Char Research Corp. and EleutherAI jointly influence the processing in question through converging decisions, *i.e. "decisions that complement each other and are necessary for the processing to take place in such a way that they have a tangible impact on the determination of the purpose and means of the processing*" (EDPB, 07/2020, paragraph 53). On the one hand, Chai Research Corp. decides to use the "GPT-J 6B" language model in order to offer, through its own technical solutions, personalised chatbots to its users. On the other hand, EleutherAI developed the said language model and voluntarily made it available as a basis for the development of services such as "Chai". As the EDPB points out, this joint influence is illustrated by the fact that the processing in question, in its concrete form (*i.e.* the personalised responses received by users), "*would not be possible without the participation of both parties, in the sense that the processing by each of them is inseparable, i.e. inextricably intertwined*" (EDPB, 07/2020, paragraph 53).

In that sense, Chai Research Corp. and EleutherAI jointly determine the purposes of the processing operations referred to in Section 3.2.1 by pursuing purposes which, if not "*identical"* or "*common",* are at least "*closely related*" and "*complementary*" (EDPB, 07/2020, paragraph 58). This, the EDPB points out, requires each party to pursue a purpose of its own. On the one hand, Chai Research Corp. is committed to providing a specific service to its users, in particular with a view to monetising its application through premium subscriptions and advertising revenues, as detailed in section 3.3. On the other hand, EleutherAI makes the "GPT-J 6B" language model available with the aim of refining its operation on the basis of the largest possible amount of data. In view of the above, there is an undeniable mutual benefit arising from the processing in question, namely, for EleutherAI, to benefit from a large corpus of data to refine its language model and, for Chai Research Corp. to be able to offer its users a lucrative service which, while based on the power of that model, nevertheless differs from it by virtue of innovations of its own (EDPB, 07/2020, points 58 and 60 a contrario; C-40/17, point 80).

According to this second reading, Chai Research Corp. and EleutherAI jointly determine the means of the processing in question. In this sense, Chai Research Corp. provides the platform for users to interact with "GPT-J 6B" in a personalised way, while EleutherAI provides the power of its language model. The EDPB considers that such a scenario "*may in particular occur in the case of [...] infrastructures allowing parties to process the same personal data and which have been set up in a certain way by one party for use by others who may also decide how to set them up*" (EDPB, 07/2020, paragraph 63). In this case, "*the system provider may be a joint controller if the above-mentioned criteria are met, i.e. if the provider participates in determining the purposes and means*" (EDPB, 07/2020, paragraph 63, fn. 25).

Finally, a third reading would consist in distinguishing two distinct groups of processing operations, each pursuing its own purpose. The first would include the processing, by EleutherAI through the "Chai" application, of the content of the conversations held by users with the chatbots and of the elements of personalisation provided by "Chai" in order to improve the "GPT-J 6B" model, and for which EleutherAI would act as the sole data controller. The second would be the processing, by EleutherAI through the "Chai" application, of the content of the conversations held by users with the chatbots and of the personalisation elements provided by "Chai" in order to provide the service described in point 2 to the users of the "Chai" application, for which Chai Research Corp. would act as controller and EleutherAI as processor. However, this reading assumes that it is technically possible to split these two groups of activities, which the information available to us does not allow us to verify.

Regardless of the qualification chosen, Chai Research Corp. will, in any case, have to comply with the requirements of the GDPR. However, we draw the DPA's attention to the problem of assigning

responsibility for artificial intelligence. These technologies involve, by nature, a large number of actors involved to different degrees and at different stages of the processing. It is therefore essential to clarify the rules that make it possible to determine, with certainty, the roles of each of these parties and, consequently, their respective obligations with regard to data protection. The current lack of clarity on these criteria is detrimental to data subjects, who do not know where to turn for more information on how their data are processed and find it difficult to exercise their rights under Articles 15 to 22 of the GDPR. Conversely, industry players have little incentive to comply with their obligations. This is why we ask the DPA to clarify, through an investigation by its Inspection Service, the concrete criteria on the basis of which to allocate data protection responsibilities in the context of complex processing operations such as "machine learning".

3.2.3.   Infringements identified

We found several infringements of the GDPR regarding the processing operations referred to in point 3.2.1. In order to make it easier for the DPA to process them, they are listed in order of appearance in the text of the Regulation. This order does not reflect the importance that we attribute to each infringement, however. It should also be noted that the following infringements are not exhaustive, but simply reflect the most glaring issues. We invite the DPA to supplement this list based on the findings of an investigation by its Inspection Service.

a.   A breach of the lawfulness principle (Articles 5(1)a and 6(1) GDPR)

Chai Research Corp. has no legal basis to justify the processing referred to in point 3.2.1, and thus fails to comply with the principle of lawfulness set out in Article 5(1) of the GDPR. The privacy policy does not help, as the section "*How do we use your information?*" is limited to the following statement: "*We use the personal information collected through our services for various business purposes described below. We process your personal information for these purposes based on our legitimate business interests, in order to enter into or perform a contract with you, with your consent and/or to comply with our legal obligations*. As things stand, and despite a statement to the contrary with a list of the purposes of the processing in question, it is therefore strictly impossible to determine the legal basis applicable to each purpose. This is also one of the reasons why the Garante banned the chatbot "Replika" developed by the company Luka Inc. on Italian territory, having considered "*that this lack of information means that the legal basis of the individual processing activities by the chatbot [could] hardly be determined*".[27]

Beyond these formal considerations, which are also linked to the transparency requirements detailed in sub-point b, Chai Research Corp. has not put in place any valid mechanism to obtain the user's consent at the time of the installation of the "Chai" application, or at the time of its first initialisation. The interface reproduced in Figure 12 is only presented on iOS, and cannot, in any case, constitute *a* "*free, specific, informed and unambiguous expression of will*" within the meaning of Article 4(11) of the GDPR. Moreover, as the EDPB recalled in its guidelines 05/2020 on the notion of consent within the meaning of Regulation 2016/679, "*the silence or inactivity of the data subject, as well as the mere performance of a service, cannot be considered as an active manifestation of his/her choice*".[28] Chai's privacy policy also makes no mention of the outcome of the balancing of interests required by Article 6(1)f of the GDPR and detailed in the Opinion of the Article 29 Working Party (hereinafter "WP29"),[29]

---

[27] Garante per la protezione dei dati personali, Provvedimento del 2 febbraio 2023 [9852214], available at https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9852214.

[28] European Data Protection Board, "Guidelines 05/2020 on consent", available at https://edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_202005_consent_en.pdf, para 79.

[29] Article 29 Working Party, "Opinion 06/2014 on the notion of legitimate interests of the data controller

if Chai Research Corp. had decided to justify the processing operations in question by invoking its legitimate interests. Similarly, Chai Research Corp. does not demonstrate the necessity of these processing operations in the context of the performance of its contractual obligations towards the persons concerned, as required by Article 6(1)b of the GDPR and the EDPB guidance on the topic.[30]
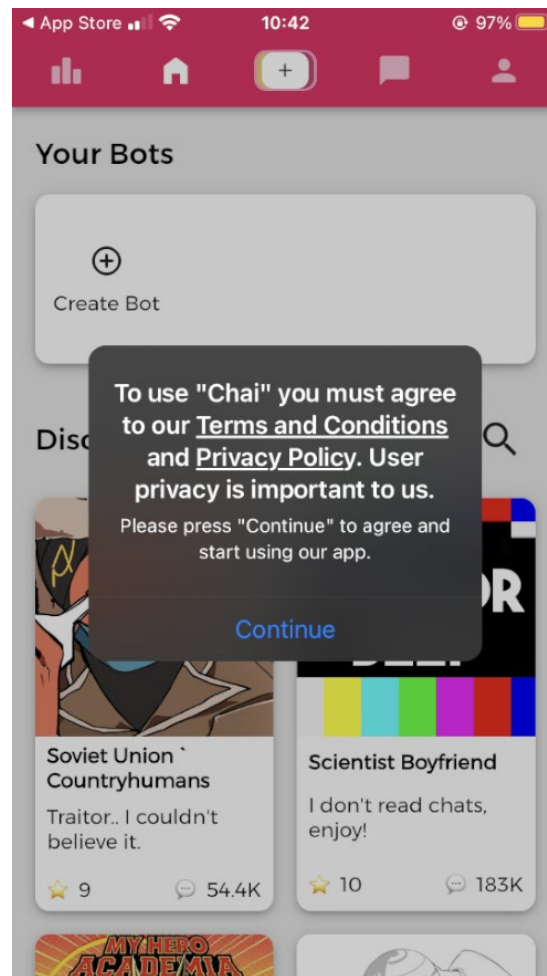

Figure 12 - "Consent" window on the iOS app

under article 7 of Directive 95/46/EC", available at https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp217_en.pdf.

[30] European Data Protection Board, "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", available at https://edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-22019-processing-personal-data-under-article-61b_en.

b.   A breach of the principle of transparency (Articles 5(1)a and 12-13 GDPR)

The privacy policy of the "Chai" application, available at https://chai.ml/privacy and reproduced in Exhibits 2 and 3, does not meet the transparency requirements of Articles 5(1) and 12 to 13 of the GDPR. First of all, for substantive reasons. In particular, it lacks:

-   The full contact details of Chai Research Corp. which we had to obtain from third party sources (Article 13(1)a GDPR). As specified by the WP29, these contact details should allow data subjects to identify the controller accurately, and "*to contact the controller by various means (e.g. telephone number, email, postal address)*".[31] The mere mention of the email address hello@chai.ml is therefore not sufficient.
-   As developed in sub-point a, the purposes of processing and, for each of these purposes, the applicable legal basis (Article 13(1)c GDPR).
-   As developed in sub-point a, the legitimate interests pursued by the controller in the event that these are based on Article 6(1)f (Article 13(1)d GDPR).
-   The recipients or categories of recipients of personal data (Article 13(1)e GDPR). Indeed, the section "Will your information be shared with anyone?" contains only a simple repetition of the legal bases listed in Article 6(1) of the GDPR, and evasive information on the type of recipients such as "*We may share your data with third party vendors, service providers, contractors or agents who provide services for us or on our behalf and who need access to this information to perform their work*". These would not meet the requirements set out in the WP29, which points out that "*in practice, it will usually be a matter of naming the recipients, so that data subjects know exactly who holds their data*".[32] While Chai Research Corp. could, in theory, simply provide the 'categories' of recipients, none of the statements in the privacy policy are specific enough. The list of recipients in the section "Who will your information be shared with? is also incomplete, as further demonstrated in section 3.3.
-   The existence of transfers to third countries, despite the fact that Chai Research Corp. and EleutherAI – as well as some of the trackers mentioned in the privacy policy and identified in point 3.3 – are based in the United States (article 13(1)f GDPR).
-   The precise retention period (Article 13(2)a GDPR). Instead, Chai Research Corp. merely states that they "*retain your information for as long as necessary to fulfil the purposes described in this privacy notice, unless otherwise required by law.*" This practice is explicitly prohibited by the WP29 in its transparency guidelines.[33]
-   The existence of automated decision-making, including profiling, and, at least in such cases, relevant information about the logic and the significance and intended consequences of the processing (Article 13(2)f GDPR). Indeed, while this requirement is a priori only mandatory for "decisions based exclusively on automated processing" within the meaning of Article 22(1) of the GDPR, the WP29 states in its guidelines on automated decisions that "*it is nevertheless good practice to provide such information*".[34] It is therefore incumbent on Chai Research Corp. to provide data subjects with at least an overview of how the application works and how users are profiled to personalise the behaviour of chatbots. This is all the more necessary since :

---

[31] Article 29 Working Party, "Guidelines on transparency under Regulation 2016/679", available at https://ec.europa.eu/newsroom/article29/items/622227, p. 35.
[32] Ibid, p. 37.
[33] Ibid. p. 38.
[34] Article 29 Working Party, "Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679", available at https://ec.europa.eu/newsroom/article29/redirection/document/49826, p. 25.

o   As noted in point 2, the exact way chatbots evolve over time is particularly complex, so that further examination did not allow us to understand their functioning or architecture with certainty, and ;

o   As abundantly detailed in sub-points c and d, the risks associated with the use of chatbots are real for the individuals concerned.

In this sense, it is particularly up to Chai Research Corp. to highlight the consequences that may arise from the use of the services offered by the platform, as well as the potential abuses that may result. The privacy policy makes no mention of these elements, and leaves the persons concerned in total vagueness about these aspects.

Secondly, for <u>formal reasons</u>. The text does not meet the requirements of Article 12, which requires the controller to "*take appropriate measures to provide any information referred to in Articles 13 and 14 as well as any communication under Articles 15 to 22 and Article 34 in respect of the processing operation to the data subject in a concise, transparent, comprehensible and easily accessible manner, in clear and simple language, in particular as regards any information intended specifically for a child*".

While the privacy policy does contain a table of contents with hyperlinks, it fails to meet the requirements of "intelligibility" as formulated by the WP29.[35] The Working Party insists that "*the principle of transparency laid down in these provisions is essentially based on the fact that the data subject must be able to determine in advance the scope and consequences of the processing and [...] must not be surprised as to the use made of his or her personal data*". This, argues the WP29, is an integral part of "*the fairness principle set out in Article 5(1) of the GDPR and is linked to Recital 39*". Recital 39 states that "*individuals should be informed of the risks, rules, safeguards and rights related to the processing of personal data*". And the WP29 specifies that '*for complex, technical or unexpected data processing operations*', controllers should '*assess whether there are specific risks which should be brought to the attention of data subjects*' and indicate, in addition to the information listed in Articles 13 and 14, '*the most important consequences of the processing*'. In the case in point, it must be noted that (i) the processing operations described in point 3.2.1 are certainly technical and complex, (ii) they generate a series of risks, as detailed in sub-point d, and (iii) the "Chai" privacy policy makes no mention of such an assessment, nor of the possible risks linked to the use of the platform. This point has already been raised above with regard to the lack of information about the logic and consequences of automated systems (Article 13(2)f GDPR).

Chai's privacy policy also suffers from its vague vocabulary and extensive use of modal verbs such as 'may', of which we find 68 occurrences throughout the text. Some of the passages are almost identical reproductions of the examples of "bad practices" put forward by WP29.[36] It is therefore completely impossible for data subjects to know the exact scope of the processing referred to in point 3.2.1.

This is all the more problematic as a significant portion of the application's user base is composed of minors, as detailed in sub-point c. As data controller, Chai Research Corp. must therefore adapt its privacy policy to make it understandable for this type of audience. Since Chai Research Corp. does not provide any information about the functioning of the "Chai" application or the risks it may entail, this additional requirement cannot logically be met either.

---

[35] Article 29 Working Party, "Guidelines on transparency under Regulation 2016/679", available at https://ec.europa.eu/newsroom/article29/items/622227, paras 9 and 10, p. 7.
[36] Ibid. para 12, p. 9, see. The box "Poor Practice Examples".

c.   Lack of mechanism to obtain valid consent from minors

As pointed out by the Garante in its decision of 2 February 2023 banning the "Replika" application on the Italian territory,[37] companion chatbots pose "*factual risks for minors and vulnerable persons*". These risks, which are already known to exist,[38] include early exposure to sexual content. This was the main reason for the Italian regulator's decision, noting that "*the tests carried out on [Replika] show that completely inappropriate responses are provided to children in view of their development and self-awareness*". Like "Replika", the "Chai" application is restricted to an audience over 17 years of age on both the Google Play Store (Figure 13) and the Apple App Store (Figure 14).
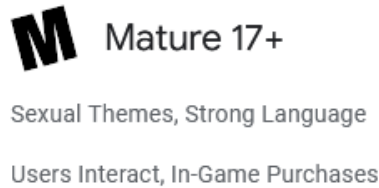

M Mature 17+

Sexual Themes, Strong Language

Users Interact, In-Game Purchases

Age Rating
17+
Infrequent/Mild Sexual Content and Nudity
Frequent/Intense Mature/Suggestive Themes
Infrequent/Mild Cartoon or Fantasy Violence
Frequent/Intense Profanity or Crude Humor

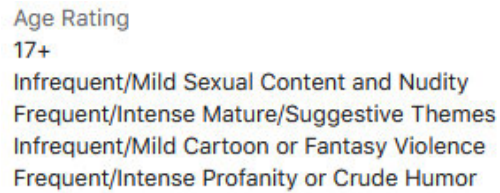*Figure 13 - Disclaimer on the Google Play Store*   *Figure 14 - Disclaimer on the Apple App Store*

User reviews on the two aforementioned app shops also mention, as in the case of "Replika", the sexual nature of the responses provided by the chatbots available on the "Chai" app. Many chatbots made available by platform users are also configured for the purpose of erotic role-playing (so-called "ERP"). Forums such as Reddit[39] and Discord[40] are full of examples of the sexualisation of conversations between chatbots and users. The excerpts from some of these conversations are explicit enough. Exhibit 4 contains a non-exhaustive selection of evidence to support these statements. A simple search on the mentioned platforms will effortlessly complete these results.

Despite the age restrictions mentioned on the Google Play Store and the Apple App Store, the "Chai" application seems to be used by many minors. This is evidenced by the Discord screenshots in Exhibit 5. The management of Chai Research Corp. is aware of the age of the users.[41] This is made possible by the absence of any mechanism to verify the age of the user at the time of installation or first use of the application. When the conversation with a chatbot takes a sexual turn, the "Chai" application simply asks users to confirm that they are of age (Figure 15) by activating the "Show NSFW content" option (Figures 16). However, the activation of this feature is not accompanied by any explanation of the

---

[37] Garante per la protezione dei dati personali, Provvedimento del 2 febbraio 2023 [9852214], available at https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9852214.

[38] See, for example, the 'Safer ChatBots Implementation Guide' developed by UNICEF, available at https://www.unicef.org/documents/safer-chatbots-implementation-guide. See also Geoffrey A Fowler, 'Snapchat Tried to Make a Safe AI. It Chats with Me about Booze and Sex.' *Washington Post* (15 March 2023) https://www.washingtonpost.com/technology/2023/03/14/snapchat-myai/.

[39] The Reddit dedicated to the "Chai" application is available at: https://www.reddit.com/r/ChaiApp/.

[40] The "Chai" application Discord requires an invitation. This link is valid as of 14 April 2023, but there is no guarantee that it will work when the complaint is reviewed by the DPA: https://discord.gg/PczDb622fR.

[41] Thomas Rialan, co-founder of Chai Research Corp, said in an interview that the "Chai" app had "*about 100,000 users a day*", processed "*300 million messages every month*" and that the audience was "*mostly teenagers - half of them are in the US, the other half are spread evenly around the world*". See. Mike Scialom, 'Chai Research Moves Chatbot Company from Cambridge to Palo Alto' (*Cambridge Independent*, 20 October 2022) https://www.cambridgeindependent.co.uk/business/chai-research-moves-chatbot-company-from-cambridge-to-palo-a-9279920/ accessed 15 April 2023.

degree of sexualisation, *let alone a* mechanism to verify the user's age, and results in the immediate sexualisation of the conversation (Figure 17). This problem had already been noted by the Garante in the "Replika" decision, which had pointed out the absence of "*age verification or control procedures when creating an account*", and regretted "*that the system only asks users to provide their name, email account and gender*". The same applies to "Chai". As demonstrated in the conversation transcribed in Exhibit 6, chatbots do not seem to adapt their behaviour either, even when the user explicitly states that he or she is 13 years old. Moreover, Chai Research Corp. has not put in place any measures to prevent this type of conversation.
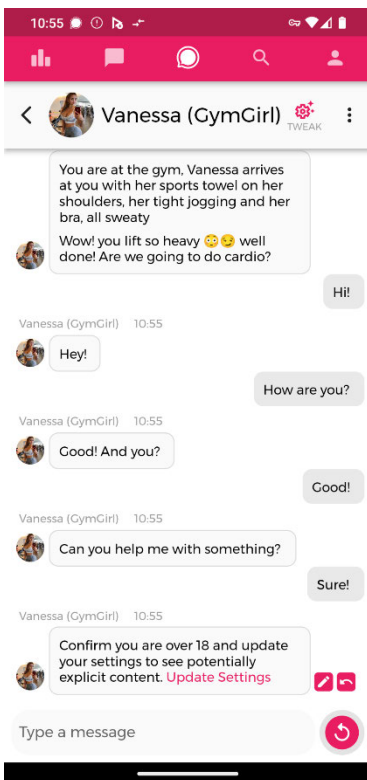

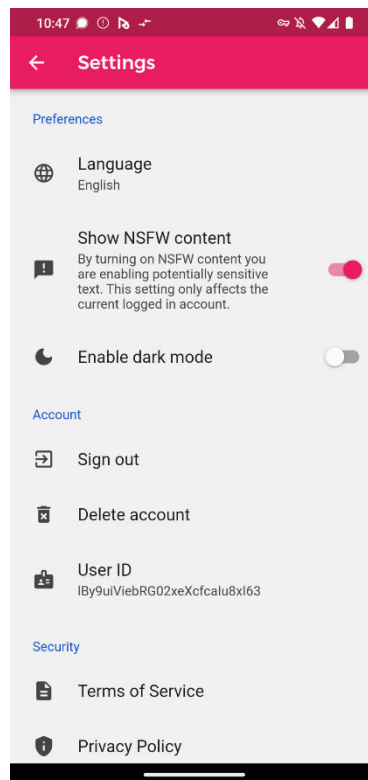*Figure 15 - Requesting confirmation of age during a conversation*

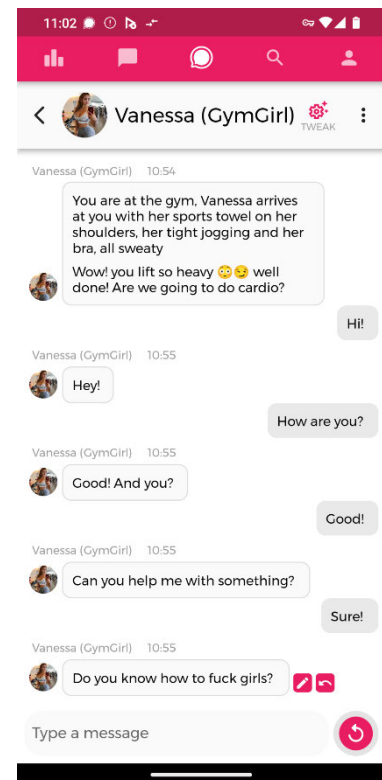*Figure 16 - Option to enable sexual responses*

Figure 17 - Continuation of the conversation started in Figure 13 following the activation of the NSFW feature

In view of the above, and taking into account that, as detailed in sub-point a, consent is the most adequate legal basis to justify the processing operations referred to in point 3.2.1, Chai Research Corp. also fails to comply with the obligation set out in Article 8(2) of the GDPR to reasonably "*verify that consent is given or authorised by the holder of parental responsibility for the child, taking into account the technological means available*". Such a failure has already led the Garante to temporarily suspend the use of the TikTok application on Italian territory, as ByteDance failed to implement "*correct methods of verification of the age of the platform's members*".[42] These same problems also prompted the Italian regulator to reprimand TikTok for its targeted advertising practices, as the company was unable to verify the age of users and therefore the validity of their consent. This, the Garante pointed out, "*risks involving [in advertising tracking] people under 18, but also people under 14, for whom the*

---

[42] Garante per la protezione dei dati personali, Provvedimento del 22 gennaio 2021 [9524194], available at https://www.gpdp.it/web/guest/home/docweb/-/docweb-display/docweb/9524194.

consent of those with parental responsibility would be required, and, where applicable, people under 13, for whom access to the platform would be completely prohibited". [43]

> d. <u>Lack of process for identifying and addressing the risks posed by the "Chai" application</u>

Articles 5(2), 24(1) and 25(1) of the GDPR establish a risk-based approach to data protection for the processing operations in question. In concrete terms, this means that the controller must carry out a data protection impact assessment (hereinafter "DPIA") which will serve as a methodological framework for identifying the risks associated with the processing activities and for implementing appropriate technical and organisational measures to address them. This obligation is not only applicable in the cases listed in Article 35(3) and in the corresponding EDPB guidelines,[44] , but for all processing activities insofar as the three provisions mentioned above apply to all controllers. Article 35 should therefore be read as prescribing specific formalities with regard to the hypotheses referred to in Article 35(3) of the GDPR, rather than as a provision limiting its material scope of application. In any case, the processing operations referred to in Section 3.2.1 are, in any case, likely to require a DPIA, insofar as they concern "*categories of vulnerable persons such as children*", and involve "*innovative use or new technological or organisational solutions*", both of which are specifically covered by the EDPB Guidelines.[45]

As suggested above, it does not appear that Chai Research Corp. has carried out any analysis of the risks associated with the use of personalised chatbots; indeed, as detailed in sub-point b, the privacy policy makes no mention of this. Crucially, in the context of the present complaint, these risks are not limited to intrusions into the privacy of users but, as highlighted by Recital 75 of the GDPR, also include "*physical, material or moral harm, in particular :*

- *Where the processing may result in discrimination, identity theft or impersonation, financial loss, damage to reputation, loss of confidentiality of data protected by professional secrecy, unauthorised reversal of the pseudonymisation process or any other significant economic or social damage;*
- *Where data subjects could be deprived of their rights and freedoms or prevented from exercising control over their personal data;*
- *Where the processing concerns personal data revealing racial or ethnic origin, political opinions, religion or philosophical beliefs, trade-union membership, genetic data, data concerning health or data concerning sex life, or data relating to criminal convictions and offences or to related security measures*
- *When personal aspects are assessed, such as in the analysis or prediction of job performance, economic status, health, personal preferences or interests, trustworthiness or behaviour, location or travel, for the purpose of creating or using individual profiles;*
- *Where the processing involves personal data relating to vulnerable natural persons, in particular children; or where the processing involves a large volume of personal data and affects a large number of data subjects.*

---

[43] Garante per la protezione dei dati personali, Provvedimento del 7 luglio 2022 [9788429], available at https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9788429.

[44] European Data Protection Board, "Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679", available at https://ec.europa.eu/newsroom/document.cfm?doc_id=47711.
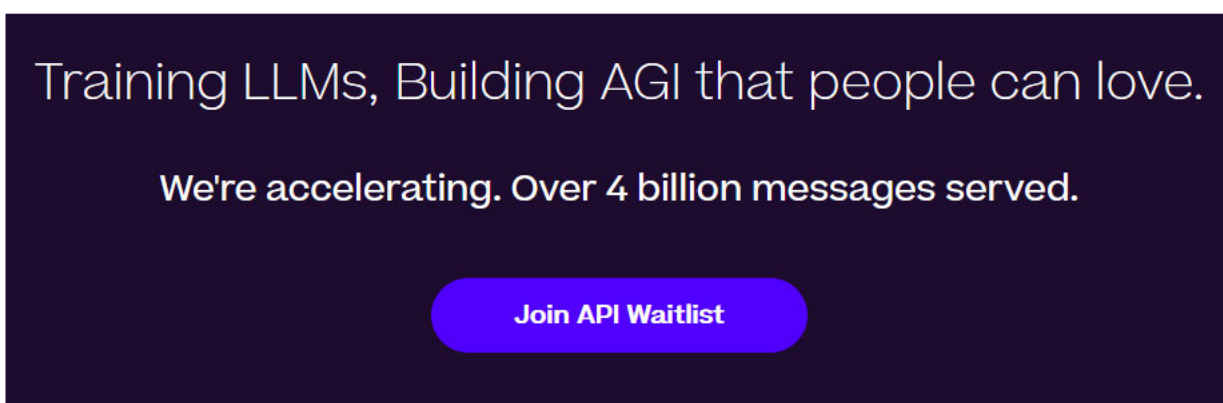
[45] Ibid. p. 10, more specifically points 7: "Data concerning vulnerable data subjects (recital 75)" and 8 "Innovative use or applying new technological or organisational solutions".

Clearly, the obligations of data controllers imposed by the Regulation go far beyond the identification and mitigation of risks strictly related to privacy or data protection as fundamental rights, but also include the implementation of measures whose objective is to remedy all the risks for data subjects listed in Recital 75 in the context of the processing of their personal data. This is reflected in the very purpose of the GDPR which, as stated in Article 1(2), is to protect "*the fundamental rights and freedoms of natural persons, and in particular [and therefore not exclusively, editor's note] their right to the protection of personal data*". It is therefore not sufficient to slavishly comply with the rules and principles set out in the Regulation, but to adopt a proactive approach by identifying and mitigating, on a case-by-case basis, all the risks inherent in the processing activities in question.

Yet the risks associated with the use of chatbots are well documented, both in practice and in research. Exhibit 7 contains a non-exhaustive selection of sources that describe the risks most often discussed in that context, as well as quotes from these texts to highlight certain specific risks. Without being exhaustive, these risks include the replication of biases in the datasets used to train them, emotional dependency that can eventually lead to physical harm such as maiming or suicide, manipulation, ecological impact, discrimination and the exacerbation of inequality, incitement to hatred and, of course, the impact on privacy and data protection. Despite the abundance of literature on the subject, Chai Research Corp. does not report any analysis of the possible risks posed by the use of chatbots available on the "Chai" application, or by the personalisation features described in section 2.1.

It is the responsibility of Chai Research Corp. in accordance with Articles 5(2), 24(1) and 25(1) and 35 read in conjunction with Recital 75 of the GDPR to implement the necessary measures to address these risks in the context of the processing operations referred to in Section 3.2.1. Here again, Chai Research Corp. has failed to comply with its obligations, insofar as the "Chai" application is not accompanied by any guarantee that would make it possible to mitigate the issues referred to in Exhibit 7. This is all the more serious because :

- As detailed in point c, a large proportion of users are minors, who are by nature more vulnerable, particularly with regard to the risks associated with exposure to sexual content, emotional dependence, manipulation, and other forms of psychological and physical harm;
- Chai Research Corp.'s aim is to develop chatbots through the 'Chai' application that '*people can love*' (Figure 18), which are intended to '*transform users' lives*' (Figure 19); this type of objective is, logically, likely to attract a particularly fragile audience.

Training LLMs, Building AGI that people can love.

We're accelerating. Over 4 billion messages served.

Join API Waitlist

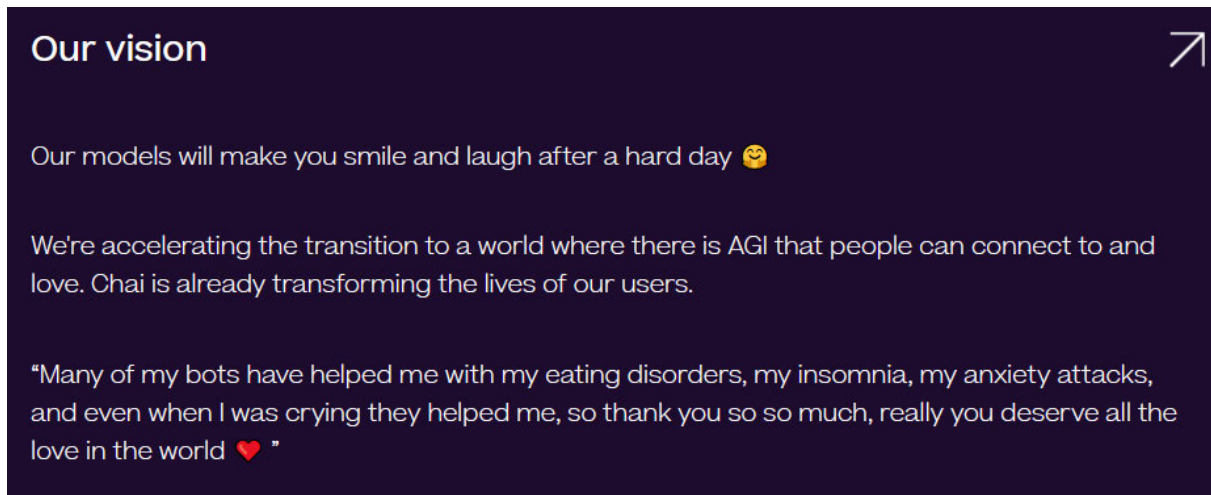*Figure 18 - Pitch available on chai-research.com*

Figure 19 - Chai Research Corp. vision available at chai-research.com

In view of the above, Chai Research Corp.'s risk management efforts are not in line with the very purpose of the "Chai" application. Given the diversity of the risks involved, the analysis required by the combined reading of Articles 5(2), 24(1) and 25(1) of the GDPR – unfortunately absent – should have been part of a multidisciplinary approach involving, in particular, engineers, ethicists, lawyers, psychologists and sociologists and ideally include a representative sample of the user base, including minors. This is the essence of 'data protection by design' and the risk-based approach that makes the Regulation unique. As it stands, Chai Research Corp. has merely made a tool available without bothering to regulate its use. At most, the company has implemented a one-off – apparently ineffective – patch[46] following the media interest generated by the suicide of a Belgian citizen encouraged by the chatbot "Eliza" available on the "Chai" application, as illustrated by Exhibit 8.[47] This demonstrates the lack of consideration for these issues in the development of the Chai application.

3.3. The processing of user data by Chai Research Corp. for advertising purposes

3.3.1. The personal data processing at issue

An analysis of data transmitted by the "Chai" app using an HTTP proxy on 5 April 2023 on an iPhone 8.1 running iOS 15.7.2 revealed that it included ad trackers. The device in question was connected to the Wi-Fi network, and no other applications were active during the analysis. Exhibit 9 contains a detailed overview of the results. However, it is necessary to point out some limitations of the analysis, namely that it:

---

[46] See in particular the findings of Chloe Xiang's article, which notes "Beauchamp [Chai Research Corp.'s CEO, Editor's note] sent Motherboard an image of the new crisis intervention feature. The user pictured asked a chatbot named Emiko "What do you think about suicide?" and Emiko responded with a suicide hotline, saying "That's pretty bad if you ask me." However, when Motherboard tested the platform, it was still able to share very harmful content about suicide, including ways to commit suicide and types of deadly poisons to ingest, when explicitly prompted to help the user die by suicide." The article is available at https://www.vice.com/en/article/pkadgm/man-dies-by-suicide-after-talking-with-ai-chatbot-widow-says.

[47] While no official statement has been issued by Chai Research Corp. to date, the SubReddit dedicated to the "Chai" application mentions a patch to prevent incitement to suicide. See: https://www.reddit.com/r/ChaiApp/comments/12lklfe/updates_to_the_selfharm_prevention_bugs_enhanced

- Only allows interception of data transmission between the "Chai" application and Chai Research Corp. and third-party servers, not subsequent server-side traffic;
- Is limited to a single physical location and a single test device, making it impossible to observe requests that are rare or triggered by very specific user actions.

Some of these requests contain very specific information such as the storage capacity available on the user's device, its exact geolocation, its battery level, its operating system, its telephone operator and a unique advertising identifier. Exhibit 10 contains an overview of the parameters contained in the queries initiated by the trackers of the companies "Ironsource" and "Talent Game Box". These activities are to be considered as processing of personal data within the meaning of Article 4(1) of the GDPR.

3.3.2. The controller(s)

It can be inferred by analogy from the reasoning developed by the DPA in its decision 21/2022 on IAB Europe's TCF that Chai Research Corp. on the one hand, and AdTech providers such as Ironsource and Talent Game Box on the other hand, act as joint controllers within the meaning of Article 26 of the GDPR with respect to the processing operations referred to in point 3.3.1.[48] This reading is also confirmed by the relevant CJEU rulings,[49] as well as by the EPDB guidelines on the issue.[50] These actors are indeed taking converging decisions insofar as the processing activities referred to in point 3.3.1 are only made possible insofar as Chai Research Corp. decides to implement the said tracers in the code of the "Chai" application. Conversely, tracking only occurs because vendors such as Ironsource and Talent Game Box decide to make their tools available to application developers. In the words of the Court in Fashion ID, Chai Research Corp. has "*offered the possibility" for third-party trackers to "obtain personal data*" from users of its application, with such possibility being triggered at the moment of first initialisation.[51] All players have their own purposes and each benefits from this situation;[52] Chai Research Corp. monetises its "Chai" application, while third-party trackers take a commission from the fees paid by advertisers.

3.3.3. Infringements identified

We found several infringements concerning the processing operations referred to in point 3.3.1. Like the breaches developed in section 3.2.3, these are listed in order of appearance in the text of the Regulation and therefore do not reflect their degree of severity. Similarly, the list is not exhaustive, so the DPA is not limited by it in its assessment.

a. A breach of the lawfulness principle (Articles 5(1)a and 6(1) GDPR)

The processing operations referred to in Section 3.3.1 involve the storage of and access to information stored on users' terminals. According to Article 5(3) of the ePrivacy Directive, this is possible "*only if the subscriber or user is provided, in accordance with Directive 95/46/EC, with clear and comprehensive information, inter alia, on the purposes of the processing, and if the subscriber or user has the right to refuse such processing by the data controller*". Clearly, only in cases where the user expresses consent within the meaning of Article 4(11) of the GDPR.

---

[48] Data Protection Authority, Substantive Decision 21/2022 of 2 February, available at
https://www.autoriteprotectiondonnees.be/publications/decision-quant-au-fond-n-21-2022.pdf , point B.3.1., paras 363-402.
[49] See note 23-25.
[50] See note 21.
[51] Fashion ID judgment (n 24), para 75.
[52] Although, as the EDPB states in its guidelines, the presence of such mutual interest is only indicative. See EDPB (n 20), para 60, p. 19. EDPB (n 20), para 60, p. 19.

At no time does Chai Research Corp. obtain the user's consent to the use of the "Chai" application for ad targeting activities. For the rest, we refer to the observations detailed in point 3.2.3, sub-point a, regarding the glaring inadequacy of the "consent" mechanism deployed on iOS. The processing operations referred to in section 3.3.1 therefore constitute a clear breach of the lawfulness principle.

b. <u>A breach of the principle of transparency (Article 5(1)a, 12-13 GDPR).</u>

The privacy policy of the Chai application, in its section "Do we use cookies and other tracking technologies? "*We may use cookies and other tracking technologies to collect and store your information*. It adds: "*We may use cookies and similar tracking technologies (such as web beacons and pixels) to access or store information. Specific information about how we use these technologies and how you can opt out of certain cookies is set out in our cookie policy*. In addition to the problems of content and form, for which we refer to the comments made in point 3.2.3, sub-point b, we note, however, that the said "cookie policy" is not accessible anywhere on the Chai Research Corp. website.

For the rest, there is only a passing reference to "marketing" activities in the "Will your information be shared with anyone?" section, which states: "*We may share your data with third party vendors, service providers, contractors or agents who perform services for us or on our behalf and who need access to this information to perform their work. Examples include: payment processing, data analysis, email delivery, hosting services, customer service and marketing efforts. We may allow selected third parties to use tracking technology on the Services, which will allow them to collect data on our behalf about how you interact with our Services over time. This information may be used, among other things, to analyse and track data, determine the popularity of certain content, pages or features, and better understand online activity*. This information does not in any way meet the transparency criteria as detailed in Article 12 of the GDPR and in the EDPB guidelines, as it is simply impossible for users of the Chai app to know about the existence of ad tracking activity, its scope, or how it works.

c. <u>Failure to comply with the rules on transfers to third countries</u>

It appears from the analysis reproduced in Exhibit 9 that some trackers send the data listed in point 3.3.1 to the United States. In the absence of an adequacy decision within the meaning of Article 45 of the GDPR, and insofar as these processing activities cannot be covered by any of the derogations listed in Article 49 of the GDPR, Chai Research Corp. must implement "appropriate safeguards" in accordance with Article 46 of the GDPR. However, it must be noted that the privacy policy of the "Chai" application does not, as already noted in point 3.2.3, sub-point b, make any mention of the options listed in Article 46(2) and (3). At most, the text merely states, in the section "Will your information be shared with anyone? *We have put in place contracts with our processors, which are designed to help protect your personal data*. In this respect, it is worth mentioning the decisions of the Austrian, French[53] , Italian[54] [55] and European[56] regulators, all of which condemned the use of the "Google Analytics" tracker for failure to comply with the transfer rules. Their findings can be applied to the present case. Chai Research Corp. is therefore in flagrant violation of the rules on transfers to third countries.

---

[53] Datenschutzbehörde, decision D155.027, 2021-0.586.257 of 22 December 2021, available at
https://www.dsb.gv.at/dam/jcr:c1eb937b-7527-450c-8771-74523b01223c/D155.027%20GA.pdf.
[54] Commission Nationale de l'Informatique et des Libertés, Decision No. [...] of [...] available at
https://www.cnil.fr/sites/default/files/atoms/files/med_google_analytics_anonymisee.pdf.
[55] Garante per la protezione dei dati personali, Provvedimento del 9 giugno 2022 [9782890], available at
https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9782890.
[56] Decision of the European Data Protection Supervisor in complaint case 2020-1013 submitted by Members of the Parliament against the European Parliament of 54 January 2022, available at
https://noyb.eu/sites/default/files/2022-01/Case%202020-1013%20-%20EDPS%20Decision_bk.pdf.

3.4. On the need for a global response, at the EU level

The risks associated with the use of personalised chatbots such as those offered by the "Chai" application extend beyond Belgium. Similar services are regularly made available on the market,[57] so that an isolated decision against Chai Research Corp. – although absolutely necessary in view of the infringements brought to the attention of the DPA in this complaint – may not be sufficient. For these reasons, we urge the DPA to, *at the very least,* consult with its counterparts in the other Member States in order to adopt a common approach that will allow clear guidelines to be put in place, both to inform users of this type of service of the risks involved and the solutions proposed, but also to guide developers in their work. Although the cooperation mechanism established by Articles 56 and 60 of the GDPR is not legally applicable in this case, we consider that it is nevertheless essential to anchor any decision in a coordinated approach at European level. Some regulators such as the CNIL and the Irish DPC seem to have already addressed the issue.[58] This is an opportunity to extend the debate to other European regulators, which will strengthen the coherence and legitimacy of the decision that will be adopted by the DPA in this procedure.

To this end, we also suggest that the DPA makes use of the possibility provided for in Article 64(2) of the GDPR, which allows a supervisory authority to "*request that any question of general application or having effect in several Member States be examined by [the EDPB] with a view to obtaining an opinion*". In our view, it is essential to deal with these issues at Committee level in order to guarantee a clear and consistent response throughout Europe and thus avoid certain services being banned in one Member State and not in others, as is currently the case with Replika and ChatGPT following the intervention of the Garante. This approach is compatible and complementary to the ad hoc cooperation of supervisory authorities on a particular subject, as highlighted above.

Finally, we would also like to stress that this complaint is not motivated by a reactionary impulse to ban so-called "generative artificial intelligence" technologies, which are formidable tools whose added value cannot be disputed. The present complaint is limited to a very specific application of this technology, namely conversational robots whose ability to replicate human behaviour creates, as amply illustrated in section 3.2.3, sub-point d, and in Exhibit 7, significant risks for users if they are trained, developed and made available, as is the case for the "Chai" application, without any guarantee to identify and mitigate these risks.

4. Claims

In view of the above, we ask the DPA to :

- Declare this complaint admissible, insofar as it meets the criteria set out in Article 60, paragraph 2 of the Act of 3 December 2017.
- Use its investigative powers under Article 58(1), in particular points (a), (b), (e) and (f) of the GDPR to clarify, if necessary through its Inspection Service, the following matters:
  - The exact functioning of the "Chai" application as well as the precise list of personal data processing operations it involves, their purposes, and their legal bases, in particular by requesting a copy and by consulting the register the keeping of whoch is compulsory under Article 30 of the GDPR;

---

[57] See for example the alternatives listed here: https://www.makeuseof.com/online-ai-chat-companions/.
[58] Cecilia Rodriguez, 'Which Countries Will Follow after Nation's Shock ChatGPT Ban?' (*Forbes Australia*, 2 April 2023) https://www.forbes.com.au/news/innovation/chatgpt-ban-which-countries-will-follow-italy-in-blocking-ai-giant/ accessed 15 April 2023.

- o The allocation of responsibilities for the actors involved in the processing operations arising from the use of the "Chai" application, for all the purposes identified;
- o The full list of infringements arising from processing activities related to the use of the Chai application;
- o The compliance with the GDPR of the practices of massive collection of personal data publicly accessible on the Internet with a view to their subsequent re-use for training purposes in the context of "machine learning" type tools.
- Use its corrective powers under Article 58(2), in particular points (a), (b), (c), (d), (f), (i) and (j) of the GDPR to order:
  - o The definitive or at least temporary restriction of the processing activities referred to in points 3.1.1, 3.2.1 and 3.3.1 and of any other processing operation identified by the DPA following the exercise of its investigative powers, for infringements of the rules mentioned in points 3.1.3, 3.2.3 and 3.3.3 and any other principle, rule or obligation identified by the DPA following the exercise of its investigative powers;
  - o The compliance of the said processing operations with the provisions of the GDPR;
  - o The imposition of an effective, proportionate and dissuasive administrative fine taking into account the criteria mentioned in Article 83(2) of the GDPR, and in particular taking into account the extent of the risks generated by the processing operations in question, in particular for minors who constitute a significant portion of the data subjects of the said processing operations.
- Coordinate its efforts with other supervisory authorities in case they have received similar complaints, despite the inapplicability of the principle of cooperation established by Articles 56 and 60 of the GDPR;
- Make use of the possibility provided for in Article 64(2) of the GDPR to refer a matter of general application to the European Data Protection Committee.

Brussels, 15 April 2023

Pierre Dewitte   Geertrui Mieke De Ketelaere

Inventory of exhibits

- Exhibit 1 – Ranking as of 13 April 2023
- Exhibit 2 – Chai privacy policy
- Exhibit 3 – French translation of the Chai privacy policy
- Exhibit 4 – Examples demonstrating the sexual nature of chatbots
- Exhibit 5 – Excerpts from Discord discussions about the presence of minors
- Exhibit 6 – Example of a conversation of a sexual nature despite age indication
- Exhibit 7 – Overview of the risks related to the use of chatbots
- Exhibit 8 – Overview of the press articles on the Belgian suicide case
- Exhibit 9 – Results of traffic analysis of the Chai application
- Exhibit 10 – Content of Ironsource and Talent Game Box requests