OVERVIEW

Overview of Joint investigation of OpenAI OpCo, LLC

Office of the Privacy Commissioner of Canada
Commission d'accès à l'information du Québec
Office of the Information and Privacy Commissioner for British Columbia
Office of the Information and Privacy Commissioner of Alberta

MAY 6, 2026

OFFICE OF THE INFORMATION & PRIVACY COMMISSIONER FOR BRITISH COLUMBIA

Overview of the Joint Investigation of OpenAI OpCo, LLC

By the Office of the Privacy Commissioner of Canada, the Commission d'accès à l'information du Québec, the Office of the Information and Privacy Commissioner for British Columbia, and the Office of the Information and Privacy Commissioner of Alberta

PIPEDA Findings #2026-002

May 6, 2026

Overview

The Office of the Privacy Commissioner of Canada ("OPC"), the Commission d'accès à l'information du Québec ("CAI"), the Office of the Information and Privacy Commissioner for British Columbia ("OIPC-BC"), and the Office of the Information and Privacy Commissioner of Alberta ("OIPC-AB"), collectively referred to as "the Offices" or "we", commenced a joint investigation into the Artificial Intelligence ("AI") research and deployment company, OpenAI OpCo, LLC ("OpenAI" or "the Respondent"). The investigation sought to examine whether OpenAI's collection, use and disclosure of the personal information of individuals in Canada, via or in relation to ChatGPT, complied with federal and provincial private sector privacy laws ("the Acts").[1]

ChatGPT was released in November 2022 and is available in Canada and globally. It is a conversational style service that can respond to users' prompts[2] and generate various types of content, such as articles or computer code. Each ChatGPT version is powered by a Large Language Model ("LLM" or "model"). LLMs are extremely large, complex machine learning systems capable of generating elaborate, plausible-sounding – but not necessarily factually accurate – content in response to queries on virtually any topic.

The investigation focused on OpenAI's GPT-3.5 and 4 models, which powered ChatGPT at the time of launching our inquiry. The Offices did not assess later models (we did, however, consider the adequacy of new measures implemented by OpenAI in response to our preliminary report of investigation), or OpenAI's other AI services (such as image or video generation). However, the findings in this report are still relevant to these products.

When referring to "the models" or "OpenAI's models" below, we mean GPT-3.5 and 4, unless otherwise specified.

---

[1] Canada's *Personal Information Protection and Electronic Documents Act* ("PIPEDA"), Quebec's *Act Respecting the Protection of Personal Information in the Private Sector* ("Quebec's Private Sector Act"), British Columbia's *Personal Information Protection Act* ("PIPA-BC") and Alberta's *Personal Information Protection Act* ("PIPA-AB").

[2] A prompt is the text input or question users provide to ChatGPT to initiate a conversation and guide its response. The prompt can also include images of what the user is looking for.

Scope of Investigation

The investigation examined OpenAI's collection (and subsequent use and disclosure) of personal information for the purpose of developing and deploying the models. OpenAI collected this information from a variety of sources, including publicly accessible Internet sources (which represent the vast majority of OpenAI's training datasets), licensed third party sources (such as specific media outlets and a large stock image vendor), and users' interactions with ChatGPT.

More specifically, the investigation sought to determine whether the company:

i. collected, used and/or disclosed personal information for purposes that a reasonable person would consider appropriate in the circumstances, and limited this collection to information which was necessary for these purposes;[3]

ii. obtained valid consent for the collection, use and disclosure of the personal information of individuals based in Canada via or in relation to ChatGPT;

iii. fulfilled its obligation to be open and transparent;

iv. took reasonable steps to ensure that the information ChatGPT generates about individuals is as accurate, complete, and up-to-date as is necessary for the purposes for which it is to be used;

v. provided individuals with the ability to obtain access to, and correct, their personal information;

vi. fulfilled its obligation to establish appropriate retention and disposal procedures for the personal information that it collects, uses and discloses; and

vii. was accountable for the personal information under its control.

To conduct this investigation, the Offices considered information from a variety of sources. In particular, the Offices relied on extensive written representations provided by the Respondent through its legal counsel and interviews with relevant OpenAI employees. This evidence included descriptions of the privacy protective measures and tools implemented by OpenAI at the various phases of its models' development and deployment, as well as the results of its internal evaluations of these measures and tools.

Technical background – Model training

To train the models, OpenAI employed a two-stage process:

i. During the initial phase, called '**pre-training**', the models gain a general understanding[4] of language by analyzing vast amounts of tokenized text data (i.e., words or parts of words converted to numerical strings) and learning to predict the next token in a sentence. Pre-training datasets are comprised of information collected from publicly accessible Internet sources and licensed third parties. At the time of developing and deploying GPT-3.5

[3] Throughout this report, the term "appropriate purpose" will be considered inclusive of "reasonable purpose" under PIPA-BC and PIPA-AB, and of "legitimate, real and important purpose" under Quebec's Private Sector Act. See s. 5(3) of PIPEDA; ss. 2, 11, 14, and 17 of PIPA-BC, SBC 2003 c 63; and ss. 2, 3, 11, 16 and 19 of PIPA-AB, SA 2003 c P-6.5. In addition, we note that s. 4 of Quebec's Private Sector Act, CQLR c P-39.1 requires that "[a]ny person carrying on an enterprise who, for a serious and legitimate reason, collects personal information on another person must determine the purposes for collecting the information before doing so."

[4] Although we use the term "understand" as a shorthand, we are mindful of the risks of anthropomorphizing AI.

and 4, OpenAI stated that it removed certain data – limited categories of websites (e.g., dark web, pornographic websites), duplicative content, and content that violated its policies – from the raw public data to reduce the processing of personal information.

ii. The '**fine-tuning**' phase seeks to further improve the model's performance on specific tasks and domains (e.g., translation, summarization, conversation) by refining its behaviour and the statistical correlations it establishes between tokens. Fine-tuning involves the use of a subset of data that is gathered through individuals' interactions with ChatGPT, as well as information provided by human trainers. OpenAI stated that at the time of developing and deploying GPT-3.5 and 4, it had a number of measures in place to mitigate privacy risks at the fine-tuning stage, including disassociating users' interactions from user accounts prior to using such data for model training and using a third-party filtering tool to remove certain categories of identifying information from these interactions. OpenAI also represented that the models were trained to refuse to provide private or sensitive information about people even if the information was publicly accessible on the open Internet and would be provided via search engines.

Analysis

The below is a summary of our findings for each issue, which are discussed in greater detail in the report.

In their analysis, the Offices balanced the need of organizations to collect, use and disclose personal information and individuals' right to privacy. Furthermore, the OPC's, OIPC-AB's and OIPC-BC's assessment of appropriate purposes was informed by the Charter values of freedom of expression and privacy.

Appropriate Purposes

The Offices accept that OpenAI's purposes for developing and deploying ChatGPT are appropriate (and "legitimate, real and important" under Quebec's Private Sector Act). However, in considering whether OpenAI complied with appropriate purposes provisions under the Acts, the Offices also considered the context of OpenAI's collection, use and disclosure of personal information collected from different sources.

Information collected from publicly accessible sources and licensed third-party datasets

Our investigation determined that the manner in which OpenAI initially collected personal information from publicly accessible websites and licensed third-party sources to train the models was overbroad and therefore inappropriate. We came to this determination considering the scale, nature, and varying levels of sensitivity of the personal information collected and used from those sources. Sources of training data included, for example, social media and discussion forums, which can contain vast amounts of personal information (including that of children), some of which will be sensitive (e.g., political views, health conditions) and potentially inaccurate (e.g., individuals' opinions or false statements about others). We found that OpenAI's mitigation measures in place at the time of training the models were not sufficient to limit the scope of its collection, use and disclosure of personal information to that which was necessary and proportional for purposes of that training. For these reasons, we determined that at the time of training the models in question, the benefits of that practice did not outweigh the risk of privacy harms. We therefore found OpenAI's collection, use and disclosure of personal information to be inappropriate in this context.

Information collected via users' interactions

The Offices accept that the collection, use and disclosure of personal information from users' interactions with ChatGPT were effective in advancing OpenAI's legitimate need to develop and deploy ChatGPT (and "necessary" within the meaning of Quebec's Private Sector Act) – in particular, to improve model outputs in responses to user prompts. We also accept that the benefits of this practice were proportional to the residual risk of privacy harm, taking into consideration the mitigation measures implemented by OpenAI at the time of training the models (including the use of a third-party tool to filter out personal information from the subset of users' interactions used to train the models), such that we found that this aspect of its training practices was not inappropriate.

Consent and Notice

The Offices found that OpenAI failed to obtain valid consent for its collection, use and disclosure of personal information for the purpose of developing and deploying the models. The Offices assessed whether OpenAI was compliant with consent provisions in its collection and use of personal information from: (i) publicly accessible websites and pages or licensed third-party sources; and (ii) users' interactions. The Offices also assessed whether OpenAI's disclosure of personal information collected from these sources, through ChatGPT's outputs, complied with consent and notice requirements.

Collection and use of personal information from publicly accessible sources

The vast majority of the information used by OpenAI to pre-train its models was obtained by crawling (automated collection, or 'scraping', of data from) publicly accessible sources (i.e. > 99%), with the remainder obtained from third-party licensed datasets.

PIPEDA, PIPA-BC and PIPA-AB have exceptions to the requirement for consent where the personal information at issue is publicly available as set out in section 7(1)(d) of PIPEDA, sections 12(1)(e), 15(1)(e) and 18(1)(e) of PIPA-BC, and sections 14(e), 17(e) and 20(j) of PIPA-AB.[5] The definition of "publicly available" is provided by each Act's regulations[6] and is distinct from a common understanding of "publicly accessible" information. While not specifically claimed by the Respondent, we note that OpenAI would not be able to rely on the exemption to consent for publicly available information given the diversity of the sources from which it collects personal information.

In any event, OpenAI took the position that it could rely on implied consent to collect and use the information from these sources to train its models.

The OPC, OIPC-AB and OIPC-BC found that OpenAI failed to obtain valid consent (implied or otherwise) for its collection and use of personal information from publicly accessible sources for model training purposes. More specifically, the three Offices noted that where the information collected and used is likely to be sensitive or where the practice is likely to be outside the reasonable expectations of the individual, express consent will generally be required. The sources in question included a wide range of personal information of varying levels of sensitivity. We did not accept that, at the time of training the models, the mitigation measures implemented by OpenAI (especially at the pre-training stage) sufficiently minimized the presence of sensitive personal information in these training datasets to allow for implied consent. Furthermore, we found that individuals (i.e., those whose information was scraped by OpenAI) would not have

---

[5] *Clearview AI v Alberta (Information and Privacy Commissioner)*, 2025 ABKB 287.
[6] Section 1 of PIPEDA's *Regulations Specifying Publicly Available Information*; Section 6 of *PIPA BC Regulations*, Prescribed source of public information and Section 7 of *PIPA AB Regulations*, Publicly available information.

reasonably expected that information posted about them publicly on the Internet could be collected and used by OpenAI to train its models, a practice which was extremely novel, and not widely understood, at the time.

The CAI concluded that OpenAI had not sufficiently documented (i) the context in which the duty to inform required under Quebec's Private Sector Act had been fulfilled with the individuals concerned by the personal information it collected or (ii) where required, the context in which the consent of these individuals had been obtained.

Collection and use of users' interactions

OpenAI asserted that it could also rely on implied consent to collect and use some personal information included in users' interactions with ChatGPT for model training purposes.

The OPC, OIPC-AB and OIPC-BC found that OpenAI should have obtained express consent for that practice. The three Offices determined that OpenAI's mitigation measures in place at the time of developing and deploying the models were not sufficient to ensure that sensitive personal information would not be included in the users' interactions used to train the models. We further found that, at the time of launching the models, OpenAI's use of users' interactions for model training fell outside of users' reasonable expectations. Indeed, many users were likely unaware or lacked basic understanding of the implications of their personal information being used for the novel practice of model training, including the potential review of their conversations by human trainers.

The CAI concluded that the information provided to some users of the models, specifically those who had an account or had downloaded the mobile application, was compliant with the specific information requirements set out under Quebec's Private Sector Act. However, the CAI found that the information provided to users of the free web version of the models was insufficient to properly inform them that their chats were being collected for model training purposes. In addition, the CAI found that, pursuant to Quebec's Private Sector Act, the privacy settings for the models should have provided, by default, that user chats would not be used for model training purposes (i.e., the most privacy-protective option).

Disclosure via ChatGPT responses

OpenAI acknowledged that, in certain circumstances, the models would disclose personal information in response to user prompts. We found that, at the time of training the models, OpenAI's training datasets would have included significant amounts of personal information of varying levels of sensitivity. While OpenAI represented that it had implemented various measures to reduce the risk that the models would disclose sensitive or private information, our investigation revealed that the categories of "sensitive or private information" that OpenAI sought to remove from model outputs were more limited than the broad range of personal information covered by the Acts (which include, for example, opinions or rumours about individuals).

The OPC, OIPC-AB and OIPC-BC therefore found that OpenAI should have obtained express consent for these disclosures where the information was either sensitive or outside the reasonable expectations of the individuals.

The CAI concluded that the same rules governing consent for the collection and use of personal information under Quebec's Private Sector Act applied to the disclosure of such personal information by OpenAI, such that it came to the same findings as it did with respect to that collection and use.

Openness (model transparency)

The OPC, OIPC-BC and OIPC-AB found that OpenAI failed to meet the openness and transparency requirements under their respective Acts.[7]

While the three Offices acknowledge that OpenAI had developed privacy communications that were generally readily accessible and written in plain language, our investigation found that certain key information was either incomplete or unclear. In particular, we determined that OpenAI was not sufficiently transparent with respect to the categories and sources of personal information that were included in its training datasets, such that individuals would not necessarily understand, by reading those communications, that information about them posted on a blog, discussion forum or social media, could be collected and used for the purpose of training OpenAI's models.

Accuracy

The OPC, OIPC-AB and OIPC-BC found that OpenAI failed to meet the accuracy requirements under the Acts.[8] [9]

LLMs are designed to produce plausible text in a conversational style, by predicting the next most likely word in a sentence, based on probabilities, context and defined parameters. However, this word may not be the most factually accurate and models can sometimes generate inaccurate, or entirely fabricated, statements. While OpenAI represented that it had carried out internal evaluations of the accuracy of its models' responses on certain topics such as math or history, our investigation found that it had not conducted an assessment to validate the general accuracy of personal information provided by the models in response to a request about, or in relation to, an individual.

Users need to understand the level of accuracy of the personal information included in ChatGPT's outputs in order to determine if that is sufficiently accurate for their intended purposes.

While OpenAI made some efforts to warn its users not to rely on the factual accuracy of outputs from its models, our testing of the models revealed that OpenAI: (i) provided insufficient notification to ChatGPT users regarding the potential for response information to be inaccurate (i.e., notice regarding the potential for inaccuracy in responses was not prominent, and more specifically, no information was made available with respect to the accuracy of personal information included in responses); (ii) did not clearly or consistently inform users of the need to verify the accuracy of facts provided; and (iii) did not consistently provide a viable mechanism for users to effectively or reliably verify those facts (i.e., sources of response content were not included in GPT-3.5 outputs, and only provided inconsistently for GPT-4 when a Browser Search feature was triggered).

Access, Correction and Deletion

The Offices found that OpenAI failed to adequately provide individuals with the ability to access, correct and delete their personal information.

Access to personal information related to a Chat GPT account or contained in the training datasets

---

[7] Regarding openness and transparency requirements, Quebec's Private Sector Act differs significantly from the other legislations discussed in this report. Consequently, the CAI did not issue any specific conclusion on this issue.
[8] Accuracy requirements being defined more narrowly under British Columbia's law, the OIPC-BC found that OpenAI contravened PIPA-BC where outputs were used to make a decision that affected the individual to whom the personal information related or were disclosed to another organization.
[9] Regarding accuracy requirements, Quebec's Private Sector Act differs significantly from the other legislations discussed in this report. Consequently, the CAI did not issue any specific conclusion on this issue.

While we accept that a well-designed self-service tool can promote efficiency in providing an individual with access to their personal information, such a tool will not generally be sufficient to meet all legal requirements. In this case, we found that the data extracts provided by OpenAI's Data Export tool to individuals seeking access to the personal information relating to their ChatGPT account were not sufficiently easy to read or understand. Furthermore, the Export Data tool would not, in each instance, provide all the personal information that OpenAI holds or discloses about a user. Finally, while the option did exist for an individual to request access beyond what was provided via the Export Data tool, OpenAI did not make this mechanism easily accessible to the requester.

With regard to personal information contained in the training datasets, our investigation found that OpenAI only provides access to personal information where it can verify that it relates directly and uniquely to the requestor. The company explained that making this connection can be an extremely complex and difficult process due to the unstructured nature of its datasets. For personal information that cannot be verified as being associated with the requestor, such as when the requestor has a common name and/or there is no way to otherwise verify attribution of the information to the requestor (e.g., through an email or telephone number), OpenAI will only tell the requestor whether their name appears in its training datasets. While we recognize that the design of OpenAI's models and the nature of the data it collects to train those models creates technical challenges, we found that in the context where OpenAI was collecting, using and disclosing vast amounts of personal information, including sensitive information, without sufficient mitigation measures to limit the information collected, it was not doing enough to comply with its access obligations under the Acts.

Correction of personal information generated by ChatGPT

With respect to requests for the correction of personal information, OpenAI indicated that if it can verify that the personal information relates to the requestor and confirm that the model did output inaccurate information, it will conduct a case-by-case assessment before attempting to implement a corrective measure. More specifically, if OpenAI cannot correct the inaccuracy due to technical challenges, it will prevent the personal information in question from appearing in ChatGPT's outputs by adding the requestor's verified personal information to a 'blocklist'.

While we acknowledge OpenAI's efforts in providing a pragmatic solution to respond to correction requests in the face of the technical challenges, we found that this approach left some gaps, most notably in instances when the company is not able to verify that the personal information relates to the requestor.

Deletion of personal information from OpenAI's models

OpenAI represented that 'untraining' or 'reverse-training' LLMs, so that they no longer use or generate specific personal information for which a deletion request has been submitted, is not currently feasible. OpenAI explained that this is because its models are trained through repeated adjustments of billions of weights (parameters) over successive runs of training datasets and do not contain or store copies of information that they 'learned' from.

OpenAI explained that it addresses requests for deletion by preventing an individual's verified personal information from appearing in ChatGPT's outputs (by adding it to a blocklist) and filtering out the information from future model training runs. OpenAI asserted that in doing this, it aims to balance privacy and data protection rights with other public interests, such as public access to information, in accordance with applicable laws.

As with requests for access and correction, OpenAI confirmed that it will only undertake the actions above when it is able to verify that the personal information uniquely relates to the requestor, which as discussed above, may often not be possible.

Retention

The Offices found that OpenAI failed to establish appropriate retention and disposal policies and procedures for the personal information that it collected, used and disclosed for the purpose of developing and deploying the models.

While OpenAI explained that it had established specific retention periods for various categories of personal information, we found that OpenAI released these models without having finalized a formal retention and deletion policy for personal information.

Furthermore, our investigation revealed that OpenAI did not have a retention schedule for the unstructured data collected from publicly accessible websites, which was stored "as long as necessary to train successive iterations of OpenAI's models."

Accountability

The Offices found that OpenAI failed to meet its accountability requirements in respect of the personal information under its control.

We recognize that OpenAI had put in place a number of structures, policies and practices to protect the personal information under its control. However, as mentioned above, we found that after having indiscriminately collected the personal information of millions of individuals in Canada and used it to train ChatGPT, without valid consent, OpenAI deployed this service without having first: (i) established the level of accuracy of personal information disclosed via model outputs – instead, it took a remedial approach to correcting systemic accuracy issues when they were discovered; and (ii) developed a retention policy for personal information collected for the purpose of developing and deploying its models.

As an indication of this lack of accountability, we note that one of OpenAI's cofounders acknowledged that the company had had concerns about ChatGPT's lack of accuracy and propensity to generate unwanted outputs when it released the tool in November 2022:[10]

> ***"Our biggest concern was around factuality, because the model likes to fabricate things.*** *But (…) other large language models are already out there, so we thought that as long as ChatGPT is better than those in terms of factuality and other issues of safety, it should be good to go. Before launch we confirmed that* ***the models did seem a bit more factual and safe than other models, according to our limited evaluations, so we decided to go ahead with the release."*** *(our emphasis added)*

[10] The inside story of how ChatGPT was built from the people who made it, MIT Technology Review, March 3, 2023.

This failure to be accountable exposed individuals to risks of harm, including breaches of their personal information, inaccuracy of information, discrimination on the basis of accurate and inaccurate information about them, in addition to other easily foreseeable individual and social harms beyond privacy that are outside of the mandate of the Offices.

OpenAI's response to our recommendations

In light of these findings, the Offices made a number of recommendations to OpenAI, with a view to allowing the development and deployment of generative AI in Canada in a sufficiently privacy-protective manner. While OpenAI generally disagreed with our findings – asserting that it was compliant with the Acts in most respects, through a combination of its existing practices and associated communications – it nonetheless engaged extensively with the Offices in an attempt to resolve the matter.

Specifically, in response to our Preliminary Report of Investigation, OpenAI informed the Offices that it had, during the course of our investigation, implemented various measures that it believed would address our recommendations. More specifically, OpenAI explained the following:

1. [Deprecation of previous models and training of new models] It has deprecated (i.e., retired) its GPT-3.5 and 4 models in July 2024 and April 2025 respectively, and used the new mitigation measures described below throughout the development and deployment of its current models powering ChatGPT. [11]
2. [Limiting use of personal information] It has developed and implemented a filtering tool to detect and mask a wide range of personal information (such as names, phone numbers, etc.) in publicly accessible Internet data and in licensed datasets used to pre-train OpenAI's models, thereby ensuring that models do not "learn" from that data. OpenAI explained that it now also uses this tool (in lieu of the previous third-party filtering tool) to redact personal identifiers from users' interactions used to fine-tune the models. This significantly reduces the amount of private and sensitive information used to train the models.
3. [Accuracy] It has introduced a new web search feature which, when activated, conducts a real-time web search and references specific web sources for the content output by ChatGPT, thereby facilitating users' independent verification of information.
4. [Accuracy] It has started proactively communicating about its assessments of the accuracy of individuals' information found in model outputs through 'model system cards' in a "Deployment Safety Hub."
5. [Accuracy] It has introduced new factuality evaluations for GPT-5, which involve asking models open-ended factual questions about people, places or concepts, or prompting them to generate biographical summaries of notable figures and evaluate the accuracy of the resulting responses.
6. [Access] It has improved the auto-response email that users receive when they submit an access request to OpenAI by email to better explain how different types of personal information can be accessed.
7. [Correction] It leverages the web search feature (see item 3 above) to facilitate the processing of correction requests. Specifically, when an individual submits a correction request, OpenAI can leverage its web search capabilities to, in response to prompts about that individual, nudge the model to conduct searches, retrieve up-

[11] Furthermore, it did not use GPT-3.5 and 4 as base models for the training of the current models.

to-date publicly accessible information from the Internet about that individual and use that information in its response.

8. [Correction and deletion] It has developed a technical solution to granularly block specific personal details about a public figure from appearing in model outputs, rather than blocking all information about that individual, thereby ensuring that ChatGPT continues to provide the public with access to information about the public figure that is of interest to them, while also ensuring that public figures can avail themselves of their privacy rights.
9. [Retention] It has implemented formal retention policies and schedules governing the retention and deletion of personal information processed in connection with ChatGPT.
10. [Retention] With respect to unstructured training data, it has implemented defined retention criteria as well as measures to ensure that deprecated and inactive datasets are no longer used in ongoing model development and are retained solely as a historical benchmark for scientific integrity purposes.

Following further discussions with the Offices, aimed at resolving our outstanding concerns, OpenAI has committed to implementing a number of additional measures:

1. [Openness and model transparency] Concurrently with the publication of this report, it will publish a bilingual Canadian blog post on its website explaining its privacy practices and take measures to promote the post and its contents in the Canadian media. The blog post will inform individuals that, among other things, users' interactions may be reviewed and used to train its models, advise users not to share sensitive information via their interactions with ChatGPT, address the question of the accuracy of its models (by adding a link to its updated "Does ChatGPT tell the truth?" article in the blog) and provide information about the categories of content used to train its models.
2. [Openness and model transparency] Within three months of the issuance of this report, it will expand its "How ChatGPT and our foundation models are developed" article to include more plain-language explanation about the sources of information used to train its models.
3. [Openness and model transparency] Within three months of the issuance of this report, it will, in the signed-out ChatGPT web experience – before the individual inputs their first user prompt – provide notice that chats may be reviewed and used to train models and advise users not to share sensitive information.
4. [Access] Within six months of the issuance of this report, it will (i) provide personal information in a more accessible and user-friendly format in its data exports, and (ii) revise the information it shares with users who are seeking a data export, to inform them about the avenues available to them if they would like to challenge the completeness, accuracy, or nature of the information provided.
5. [Retention] Within six months of the issuance of this report, with respect to future datasets collected lawfully, which are deprecated and solely used as a historical benchmark for scientific integrity purposes, it will:
    a. confirm in a report to be provided to the Offices that strong technical and organizational controls are in place to ensure that the datasets retained for related scientific integrity purposes are not used for active model development once they are no longer needed for that purpose;

b. to the extent that these retained datasets contain personal information, continue to comply with applicable data subject rights, as required by law; and

c. continue its existing process of regularly re-evaluating whether retention of each dataset remains necessary pursuant to its established criteria.

6. [Children's privacy] Within six months of the issuance of this report, it will test the addition of a protective measure for the minor family members of public figures (who are not themselves public figures), so that the models refuse requests for the name or date of birth of minor family members of such individuals, even if such information is publicly accessible through a current online citation.

7. [Reporting] It will provide the Offices with quarterly reports to confirm and demonstrate, with detailed submissions and corroborating evidence, compliance with the above commitments until all have been met.

Conclusion

OPC: Based on OpenAI's commitments, and in line with a pragmatic and flexible interpretation of PIPEDA and the necessity to balance the privacy rights of individuals with the need for businesses to use personal information for appropriate purposes, the OPC concludes that the issues under investigation are **well-founded and conditionally resolved**.

In coming to this conclusion, the OPC considered among other factors that the measures implemented, or to be implemented by OpenAI will significantly reduce the residual risk of harm to individuals associated with the collection, use and disclosure of their personal information in the development and deployment of ChatGPT models. In particular:

i. the new mitigation measures implemented by OpenAI, including the new filtering tool, will significantly limit the amount of personal information, and of sensitive personal information, included in OpenAI's training datasets;

ii. public awareness regarding AI and LLMs has evolved since the launch of ChatGPT (GPT-3.5), as have individuals' reasonable expectations with respect to how their personal information may be collected and used to train those models, and those expectations will be further informed by the additional transparency measures to be implemented by OpenAI; and

iii. OpenAI's deprecation of previous models trained without these protective measures, effectively ceases its overbroad use of personal information for the purposes of training ChatGPT models.

AB and BC: The OIPC-BC and OIPC-AB similarly took a pragmatic and flexible approach to interpreting the respective legislation, as is consistent with the modern approach, but these statutes are, in certain key areas, more specific and explicit than PIPEDA. In particular, these statutes meet the standard set in PIPEDA related to appropriate purpose and thus are substantially similar but are more specific than PIPEDA. For this reason, the OIPC-BC and OIPC-AB did not have the latitude to interpret the statutes as broadly as the OPC did. The OIPC-BC and OIPC-AB find that OpenAI's models are based on scraped data for which OpenAI has not obtained, and cannot obtain, consent under PIPA-BC and PIPA-AB. While the OIPC-AB and OIPC-BC are encouraged by the new measures aimed at compliance taken by OpenAI since this investigation has been initiated and those which it has further committed to implement, these are not sufficient to meet the foundational requirement for consent in PIPA-BC and PIPA-AB. Despite this finding, the OIPC-BC and OIPC-AB joined

the OPC and CAI in making joint recommendations and in monitoring implementation of the measures to which OpenAI has committed.

CAI: The CAI considers Issues 1, 5 and 7 (i.e., appropriate purposes, individual rights and accountability) to be well-founded and conditionally resolved, and Issues 2 (consent) and 6 (retention) to be well-founded and unresolved. Given the specificities of its legislation, the CAI does not issue a finding on Issues 3 and 4 (i.e., openness and accuracy). Furthermore, the CAI has made specific recommendations with respect to consent and retention to bring OpenAI in compliance with Quebec's Private Sector Act. The CAI intends to monitor OpenAI's implementation of the joint recommendations, as well as its own specific recommendations. The CAI will take this into consideration in its assessment of whether to undertake any additional verification or investigative action and / or issue any further recommendations or orders related to the compliance of OpenAI's practices with Quebec's Private Sector Act.

More generally, the Offices expect that OpenAI will continue to effectively implement and improve its mitigation measures and develop further innovative techniques in the future to maintain and improve privacy protection in the development and deployment of its models.

Finally, while this report aims at addressing and mitigating the risk to privacy associated with the development and deployment of LLMs, we recognize that this technology raises many other questions and challenges, including societal and ethical ones, which regulators, academics and courts around the world are currently trying to assess and address. We trust that this collective effort will contribute to further shaping and defining a robust framework for the future development of Generative AI, in Canada and around the world.